

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: September 24, 2004, 11:06:56 ; Search time 21.0943 Seconds
(without alignments)
3126.121 Million cell updates/sec

Title: US-10-658-989a-1

Perfect score: 1184
Sequence: 1 GPCEPPTGLPGPGERGG.....GEQGVPGDLGAPGSPAGG 209

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1017041 seqs, 315518202 residues

Total number of hits satisfying chosen parameters: 1017041

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

SPTREMBL 25:*

- 1: sp_archaea:*
- 2: sp_bacteria:*
- 3: sp_fungi:*
- 4: sp_human:*
- 5: sp_invertebrate:*
- 6: sp_mammal:*
- 7: sp_mmc:*
- 8: sp_organalle:*
- 9: sp_phase:*
- 10: sp_plant:*
- 11: sp_rodent:*
- 12: sp_virus:*
- 13: sp_vertebrate:*
- 14: sp_unclassified:*
- 15: sp_rvirus:*
- 16: sp_bacteriap:*
- 17: sp_archaeap:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1149	97.0	1461	4	O76045 homo sapien
2	1149	97.0	1464	4	Q8N473 homo sapien
3	1122	94.8	1453	11	Q63079 rattus norv
4	1116	94.3	1453	11	Q81079 mus musculu
5	989	83.5	1450	13	Q9Y1B4 cynops pyrr
6	981	82.9	1445	13	Q3251 rana catesb
7	962.5	81.3	1449	13	Q802B5 xenopus lae
8	941	79.5	1447	13	Q91B91 xenopus lae
9	860	72.6	1418	6	Q28396 equus cabal
10	860	72.6	1487	6	Q77753 canis famil
11	859	72.6	1160	4	Q14046 homo sapien
12	859	72.6	1269	13	Q7C227 gallus gall
13	859	72.6	1420	13	Q90W37 gallus gall
14	859	72.6	1487	4	Q14047 homo sapien
15	850	71.8	1486	13	Q91717 xenopus lae
16	850	71.8	1486	13	Q7ZT16 xenopus lae

17	844	71.3	1491	13	Q91718
18	844	71.3	1491	13	Q7ZTM4
19	839	70.9	1419	11	Q63123
20	839	70.9	1442	11	Q82031
21	839	70.9	1442	11	Q62033
22	839	70.9	1459	11	Q62032
23	835	70.5	1419	11	Q80X38
24	835	70.5	1419	11	Q80VY3
25	833	70.4	1418	13	Q9W7B9
26	833	70.4	1449	13	Q910C0
27	798	67.4	1347	4	Q86QB3
28	776	65.5	1138	4	Q9UML6
29	776	65.5	1163	4	Q8N6U4
30	771	65.1	1497	11	Q7TMS0
31	768	64.9	1458	13	Q910B9
32	765	64.6	1497	11	Q61431
33	763	64.4	998	11	Q8CFM4
34	763	64.4	1222	11	Q8K173
35	763	64.4	1464	11	Q8BK12
36	763	64.4	1464	11	Q8BLW4
37	763	64.4	1464	11	Q7T732
38	760	64.2	1314	11	Q8CGA5
39	745	62.9	1366	4	Q15177
40	745	62.9	1366	4	Q725S6
41	717.5	60.6	1258	13	Q8AW11
42	705.5	59.6	1352	13	Q90Y00
43	691.5	58.4	1352	13	Q8U0J4
44	689.5	58.2	1346	13	Q8U0J3
45	678.5	57.3	1346	13	Q801M5

ALIGNMENTS

RESULT 1

O76045 PRELIMINARY; PRT; 1461 AA.

AC O76045; 1998 (TREMREL. 08, Created)
DT 01-NOV-1998 (TREMREL. 12, Last sequence update)
DT 01-OCT-2003 (TREMREL. 25, Last annotation update)
DE Pro alpha 1(I) collagen.
GN COL1A1
OS Homo sapiens (Human)
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
CX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=851130970; PubMed=2857713;
RA Chu M.L., de Wet W., Bernard M., Ramirez F.;
RT "Fine structural analysis of the human pro-alpha 1 (I) collagen gene.
Promoter structure, Alu repeats, and polymorphic transcripts.";
J. Biol. Chem. 260:2315-2320(1985).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=88329734; PubMed=2843432;
RA D'Alessio M., Bernard M., Pretorius P.J., de Wet W., Ramirez F.;
RT "Complete nucleotide sequence of the region encompassing the first
twenty-five exons of the human pro alpha 1(I) collagen gene.";
Gene 67:105-115(1988).
RN [3]
RP SEQUENCE FROM N.A.
RX MEDLINE=89025644; PubMed=3178743;
RA Tromp G., Kuivaniemi H., Stacey A., Shikata H., Baldwin C.T.,
Jenisch R., Prockop D.J.;
RT "Structure of a full-length cDNA clone for the prepro alpha 1(I) chain
of human type I procollagen.";
Biochem. J. 253:919-922(1988).
RN [4]
RP SEQUENCE FROM N.A.
RX MEDLINE=91138770; PubMed=1995349;
RA Maatta A., Bornstein P., Penttinen R.P.;

RT "Highly conserved sequences in the 3'-untranslated region of the
RT COL1A1 gene bind cell-specific nuclear proteins.";
RL FEBS Lett. 279:9-13(1991).
RN [5]
RP SEQUENCE FROM N.A.
RX MEDLINE=92157916; PubMed=1787829;
RA Westerhausen A., Constantinou C.D., Pack M., Peng M.Z., Hanning C.,
RA Olsen A.S., Prockop D.J.;
RT "Completion of the last half of the structure of the human gene for
RT the Pro alpha 1 (I) chain of type I procollagen (COL1A1).";
RL Matrix 11:375-379(1991).
RN [6]
RP SEQUENCE FROM N.A.
RX MEDLINE=98107942; PubMed=9443852;
RA Korkko J., Ala-Kokko L., De Paeppe A., Nuytinck L., Earley J.,
RA Prockop D.J.;
RT "Analysis of the COL1A1 and COL1A2 genes by PCR amplification and
RT scanning by conformation-sensitive gel electrophoresis identifies only
RT COL1A1 mutations in 15 patients with osteogenesis imperfecta type I;
RT identification of common sequences of null-allele mutations.";
RL Am. J. Hum. Genet. 62:98-110(1998).
RN [7]
RP SEQUENCE FROM N.A.
RA Korkko J.M., Earley J.J., Nuytinck L., DePaeppe A., Prockop D.J.,
RA Ala-Kokko L.;
RL Submitted (MAY-1999) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF017178; AAB94054.2; -
DR GO; GO:0005581; C:collagen; IEA.
DR GO; GO:0005201; F:extracellular matrix structural constituent; IEA.
DR InterPro; IPR008161; Clg helix.
DR InterPro; IPR008160; Collagen.
DR InterPro; IPR002181; Fibrinogen_C.
DR InterPro; IPR000885; Fib_collagen_C.
DR InterPro; IPR001007; VWF_C.
DR Pfam; PF01410; COLFI; 1.
DR ProDom; PD000007; Clg helix; 2.
DR ProDom; PD002078; Fib_collagen_C; 1.
DR SMART; SM00038; COLFI; 1.
DR SMART; SM00214; VMC; 1.
DR PROSITE; PS01208; VWFC_1; 1.
DR PROSITE; PS0184; VWFC_2; 1.
KW Hypothetical protein; Collagen.
SQ SEQUENCE 1461 AA; 138630 MW; 9ACF6DE30EA78E21 CRC64;

Query Match 97.0%; Score 1149; DB 4; Length 1461;
Best Local Similarity 93.6%; Pred. No. 8.1e-72;
Matches 206; Conservative 0; Mismatches 2; Indels 12; Gaps 1;
QY 1 GPP-----GPPGPTGLPGRGSGRGPAGDGVAGPKGPPAGERSGPGA 48
Db 455 GPPGAGEGKRGARGPPTGLPGRGSGRGPAGDGVAGPKGPPAGERSGPGA 514
QY 49 GPKGSGEAGRPGAGLPGAKLGTSGSPSPDGTGTPGAGQDGRPGPPPGARGQA 108
Db 515 GPKGSGEAGRPGAGLPGAKLGTSGSPSPDGTGTPGAGQDGRPGPPPGARGQA 574
QY 109 GVMGFPKGAAGRPGKAGRGVPPGAVGPKGCGEAGAGQPPGPPGAGRGEGGPA 168
Db 575 GVMGFPKGAAGRPGKAGRGVPPGAVGPKGCGEAGAGQPPGPPGAGRGEGGPA 634
QY 169 GSPGFQGLPGPAGPPGAGKPGEGQVPGDLGAPGSPGAG 208
Db 635 GSPGFQGLPGPAGPPGAGKPGEGQVPGDLGAPGSPGARG 674

RESULT 2
Q8N473
ID Q8N473 PRELIMINARY; PRT; 1464 AA.
AC Q8N473
DT 01-OCT-2002 (TrEMBLrel. 22, Created)
DT 01-OCT-2002 (TrEMBLrel. 22, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)

DE Hypothetical protein.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Brain;
RA Strausberg R.;
RL Submitted (AUG-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; BC036531; AAB36531.1; -
DR GO; GO:0005581; C:collagen; IEA.
DR GO; GO:0005201; F:extracellular matrix structural constituent; IEA.
DR InterPro; IPR008161; Clg helix.
DR InterPro; IPR008160; Collagen.
DR InterPro; IPR002181; Fibrinogen_C.
DR InterPro; IPR000885; Fib_collagen_C.
DR InterPro; IPR001007; VWF_C.
DR Pfam; PF01410; COLFI; 1.
DR Pfam; PF01391; Collagen; 18.
DR ProDom; PD000007; Clg helix; 2.
DR ProDom; PD002078; Fib_collagen_C; 1.
DR SMART; SM00038; COLFI; 1.
DR SMART; SM00214; VMC; 1.
DR PROSITE; PS01208; VWFC_1; 1.
DR PROSITE; PS0184; VWFC_2; 1.
KW Hypothetical protein; Collagen.
SQ SEQUENCE 1464 AA; 139011 MW; B0581F8D1C89DDE8 CRC64;
Query Match 97.0%; Score 1149; DB 4; Length 1464;
Best Local Similarity 93.6%; Pred. No. 8.1e-72;
Matches 206; Conservative 0; Mismatches 2; Indels 12; Gaps 1;
QY 1 GPP-----GPPGPTGLPGRGSGRGPAGDGVAGPKGPPAGERSGPGA 48
Db 458 GPPGAGEGKRGARGPPTGLPGRGSGRGPAGDGVAGPKGPPAGERSGPGA 517
QY 49 GPKGSGEAGRPGAGLPGAKLGTSGSPSPDGTGTPGAGQDGRPGPPPGARGQA 108
Db 518 GPKGSGEAGRPGAGLPGAKLGTSGSPSPDGTGTPGAGQDGRPGPPPGARGQA 577
QY 109 GVMGFPKGAAGRPGKAGRGVPPGAVGPKGCGEAGAGQPPGPPGAGRGEGGPA 168
Db 578 GVMGFPKGAAGRPGKAGRGVPPGAVGPKGCGEAGAGQPPGPPGAGRGEGGPA 637
QY 169 GSPGFQGLPGPAGPPGAGKPGEGQVPGDLGAPGSPGAG 208
Db 638 GSPGFQGLPGPAGPPGAGKPGEGQVPGDLGAPGSPGARG 677
RESULT 3
Q63079
ID Q63079 PRELIMINARY; PRT; 1453 AA.
AC Q63079
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-JUN-1998 (TrEMBLrel. 06, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Collagen alpha1 (Fragment).
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Rattus.
OX NCBI_TaxID=10116;
RN [1]
RP SEQUENCE OF 1-1092 FROM N.A.
RC STRAIN=Sprague-Dawley; TISSUE=Tooth;
RA Brandsten C., Lundmark C., Christerson C., Hammarstrom L., Wurtz T.;
RT "Expression of Collagen alpha1(I) mRNA variants during Tooth and Bone
RT Formation in the Rat.";
RL J. Dent. Res. 0:0-0(0).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=Sprague-Dawley; TISSUE=Tooth;
RA Wurtz T.;

```
RL Submitted (FEB-1998) to the EMBL/GenBank/DBJ databases.
DR EMBL; Z78279; CAB01633.1; -.
DR GO; GO:0005581; C:collagen; IEA.
DR GO; GO:0005201; F:extracellular matrix structural constituent; IEA.
DR InterPro; IPR008161; Clg_helix.
DR InterPro; IPR008160; Collagen.
DR InterPro; IPR002181; Fibrinogen_C.
DR InterPro; IPR000885; Fib_collagen_C.
DR InterPro; IPR001007; VWFC.
DR Pfam; PF01410; COLFI; 1.
DR Pfam; PF01391; Collagen; 18.
DR ProDom; PD000007; Clg_helix; 3.
DR ProDom; PD002078; Fib_collagen_C; 1.
DR SMART; SM00038; COLFI; 1.
DR SMART; SM00214; VWC; 1.
DR ProSite; PS01208; VWFC_1; 1.
DR ProSite; PS01084; VWFC_2; 1.
KW Collagen.
FT NON TER
SQ SEQUENCE 1453 AA; 137887 MW; E6896BDC19A41D8 CRC64;

Query Match 94.8%; Score 1122; DB 11; Length 1453;
Best Local Similarity 91.4%; Pred. No. 5.9e-70;
Matches 201; Conservative 2; Mismatches 5; Indels 12; Gaps 1;

QY 1 GPP-----GEPGPTGLPGPPGERGSGRPGPGADGVAGPKGPGAGERSGPGA 48
DB 447 GPPGAGEEKGARGEPGSGLPGLPGPPGERGSGRPGPGADGVAGPKGSGERSGPGA 506

QY 49 GPKSGPGEAGRPGEAGLPGAKGLTSGPSGPDGKTGPPGAGQDGRPGPPGARGQA 108
DB 507 GPKSGPGEAGRPGEAGLPGAKGLTSGPSGPDGKTGPPGAGQDGRPGPPGARGQA 566

QY 109 GVMGPPGKGAAGEPKKAGRGVPPGAVGPAKDGAGAGQGPAGPAGERGEGQA 168
DB 567 GVMGPPGKGTAGEPKAGRGVPPGAVGPAKDGAGAGQGPAGPAGERGEGQA 626

QY 169 GSPGFQGLPGPAGPGEAGKPGEGQVPGDLGAPGSPGAG 208
DB 627 GSPGFQGLPGPAGPGEAGKPGEGQVPGDLGAPGSPGARG 666

RESULT 4
Q810J9 PRELIMINARY; PRT; 1453 AA.
AC Q810J9
DT 01-JUN-2003 (TrEMBLrel. 24, Created)
DT 01-JUN-2003 (TrEMBLrel. 24, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Hypothetical protein.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sclurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=FVB/N; TISSUE=Colon;
RA Strausberg R.;
RL Submitted (APR-2003) to the EMBL/GenBank/DBJ databases.
DR EMBL; SC050014; RAH50014.1; -.
DR GO; GO:0005581; F:extracellular matrix structural constituent; IEA.
DR InterPro; IPR008160; Collagen.
DR InterPro; IPR002181; Fibrinogen_C.
DR InterPro; IPR000885; Fib_collagen_C.
DR Pfam; PF01410; COLFI; 1.
DR Pfam; PF01391; Collagen; 18.
DR ProDom; PD002078; Fib_collagen_C; 1.
DR SMART; SM00038; COLFI; 1.
DR SMART; SM00214; VWC; 1.
DR ProSite; PS01208; VWFC_1; 1.
DR ProSite; PS01084; VWFC_2; 1.

QY 169 GSPGFQGLPGPAGPGEAGKPGEGQVPGDLGAPGSPGAG 208
DB 627 GSPGFQGLPGPAGPGEAGKPGEGQVPGDLGAPGSPGARG 666

RESULT 5
Q9YIB4 PRELIMINARY; PRT; 1450 AA.
AC Q9YIB4
DT 01-MAY-1999 (TrEMBLrel. 10, Created)
DT 01-MAY-1999 (TrEMBLrel. 10, Last sequence update)
DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE Alpha 1 type I collagen.
OS Cynops pyrrhogaster (Japanese common newt).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Amphibia; Batrachia; Caudata; Salamandridae; Salamandridae; Cynops.
OX NCBI_TaxID=8330;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Regenerate forelimbs;
RX MEDLINE=99407244; PubMed=10474166;
RA Asahina K.; Ohara M.; Yoshizato K.;
RT "Expression of genes of type I and type II collagen in the formation
RT and development of the blastema of regenerating newt limb.";
RL Dev. Dyn. 216:59-71(1999).
DR EMBL; AB015438; BAA36973.1; -.
DR GO; GO:0005581; C:collagen; IEA.
DR GO; GO:0005201; F:extracellular matrix structural constituent; IEA.
DR InterPro; IPR008161; Clg_helix.
DR InterPro; IPR008160; Collagen.
DR InterPro; IPR000885; Fib_collagen_C.
DR InterPro; IPR001007; VWFC.
DR Pfam; PF01410; COLFI; 1.
DR Pfam; PF01391; Collagen; 18.
DR ProDom; PD000007; Clg_helix; 1.
DR ProDom; PD002078; Fib_collagen_C; 1.
DR SMART; SM00038; COLFI; 1.
DR SMART; SM00214; VWC; 1.
DR ProSite; PS01208; VWFC_1; 1.
DR ProSite; PS01084; VWFC_2; 1.
KW Collagen.
SQ SEQUENCE 1450 AA; 137564 MW; ABF8A74841B87B7C CRC64;

Query Match 83.5%; Score 989; DB 13; Length 1450;
Best Local Similarity 80.5%; Pred. No. 9.5e-61;
Matches 177; Conservative 7; Mismatches 24; Indels 12; Gaps 1;

QY 1 GPP-----GEPGPTGLPGPPGERGSGRPGPGADGVAGPKGPGAGERSGPGA 48
DB 444 GPPGAGEEKGARGEPGSGLPGLPGPPGERGSGRPGPGADGVAGPKGPGAGERSGPGA 503

QY 49 GPKSGPGEAGRPGEAGLPGAKGLTSGPSGPDGKTGPPGAGQDGRPGPPGARGQA 108
DB 504 GPKSGTGESRPGEPGLPGAKGLTSGPSGPDGKTGPGAGAGQDGHPPGPGSGARGOS 563
```

AC	Q802B5;
DT	01-JUN-2003 (TrEMBLrel. 24, Created)
DT	01-JUN-2003 (TrEMBLrel. 24, Last sequence update)
DT	01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE	Hypothetical protein.
OS	Xenopus laevis (African clawed frog).
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC	Amphibia; Batrachia; Anura; Mesobatrachia; Pipoidae; Pipidae;
OC	Xenopodinae; Xenopus.
OX	NCBI_TaxID=8355;
RN	[1]
RP	SEQUENCE FROM N.A.
RC	TISSUE=Embryo;
RA	Klein S., Strausberg R.;
RL	Submitted (MAR-2003), to the ENBL/GenBank/DBJ databases.
DR	ENBL; BC049829; AAA49829.1; -.
DR	GO; GO:0005581; C:collagen; IBA.
DR	GO; GO:0005201; F:extracellular matrix structural constituent; IEA.
DR	InterPro; IPR008160; Collagen.
DR	InterPro; IPR000885; Fib collagen_C.
DR	InterPro; IPR001007; VWF_C.
DR	Pfam; PF01410; COLFI; 1.
DR	Pfam; PF01391; Collagen; 18.
DR	ProDom; PD002078; Fib_collagen_C; 1.
DR	SMART; SM00038; COLFI; 1.
DR	SMART; SM00214; VWC; 1.
DR	PROSITE; PS01208; VWFC_1; 1.
DR	PROSITE; PS01184; VWFC_2; 1.
_KW	Hypothetical protein.
SQ	SEQUENCE 1449 AA; 137464 MW; 4B32E20BECCB978C CRC64;
Query Match	
Best Local Similarity 79.5%; Score 962.5; DB 13; Length 1449;	
Matches 175; Conservative 8; Mismatches 22; Indels 15; Gaps 2	
QY	1 GPP-----GPGPTGLPGPPGERGGSGRFGPCADVAGPKPAGERGSPGA 48
Db	444 GPPFSGEGEKRGSGRFGFPAGFPGPAGERGGSGRFGPCSDASGKGPAGERGFVGPA 503
QY	49 GKPSGPGAGRPGEAGLPKAAGLTGSPGSPDGKTGPGPAGDGRPGPPPGARGQA 108
Db	504 GKSSGESGRPGFGLPKAGLGTGSPGSPDGKTGPGAPQDGRAGPPPGARGQS 563
QY	109 GWMGFPKGAAAGPPKAGRGVPGPPGAYGVPAGKDGEAGAOGPPGAGPAGERGQPGA 168
Db	564 GWMGFPKGAAAGPPKNGEKGVAGPPGAVGLPKGDGAGAGPPGPFAGPAGERGQPGA 623
QY	169 GSPFCQLPGPAGPPGAGKEGQGVPGDLGAPSGSPAG 208
Db	624 GSPFCQLPGPAGSGKPGQGVPDV---GPSGPAG 660
RESULT 8	
Q9IB91 PRELIMINARY; PRT; 1447 AA.	
ID	Q9IB91
AC	Q9IB91;
DT	01-OCT-2000 (TrEMBLrel. 15, Created)
DT	01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT	01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE	Type I collagen alpha 1.
GN	COL1A1
OS	Xenopus laevis (African clawed frog).
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC	Amphibia; Batrachia; Anura; Mesobatrachia; Pipoidae; Pipidae;
OC	Xenopodinae; Xenopus.
OX	NCBI_TaxID=8355;
RN	[1]
RP	SEQUENCE FROM N.A.
RA	Goto T., Katada T., Kinoshita T., Kubota H.Y.;
RT	"Expression and characterization of Xenopus type I collagen alpha 1 (COL1A1) during embryonic development."
RL	Submitted (NOV-1999); to the ENBL/GenBank/DBJ databases.
DR	EMBL; AB034701; BAA94972.1; -.

DR GO; GO:0005581; C:collagen; IEA.
 DR GO; GO:0005201; F:extracellular matrix structural constituent; IEA.
 DR InterPro; IPR008161; Clg helix.
 DR InterPro; IPR008160; Collagen.
 DR InterPro; IPR000885; Fib_collagen_C.
 DR InterPro; IPR001007; VWF_C.
 DR Pfam; PF01410; COLFI; 1.
 DR Pfam; PF01391; Collagen; 18.
 DR Pfam; PF00093; vwc; 1.
 DR ProDom; PD000007; Clg helix; 2.
 DR ProDom; PD002078; Fib_collagen_C; 1.
 DR SMART; SM00038; COLFI; 1.
 DR SMART; SM00214; VWC; 1.
 DR PROSITE; PS01208; VWF_C_1; 1.
 DR PROSITE; PS01184; VWF_C_2; 1.
 KW Collagen.
 SQ SEQUENCE 1447 AA; 137446 MW; A4AGDD2B4158E38B CRC64;
 Query Match 79.5%; Score 941; DB 13; Length 1447;
 Best Local Similarity 78.5%; Pred. No. 2e-57;
 Matches 172; Conservative 8; Mismatches 25; Indels 14; Gaps 2;
 QY 1 GPPGPGP-----PTGLPGPPGGRGGSGRFGPGADGVAGPKGAPGGERGSGPGAG 49
 Db 444 GPPGSGEGKSGRGEPPAGPPGAPGERGGSGRFGSGDGASGPKGAPGERGPGVGPAG 503
 QY 50 PKGSGEAGRGEAGLPKAGLTGSPGSGPDGKTGPPGAGQDGRGPPGPPGARGQAG 109
 Db 504 PKGSSGESRGEPGLPKAGLTGSPGSGPDGKTGPGAGPQDGRAGPPGPPGARGQSG 563
 QY 110 VMGPPGPKGAGEPKGAGRGVPGPGAVGAGKDGAGAGPQDGRGPPGAPGAGERGEGQPGAG 169
 Db 564 VMGFFPKGAGEPKGAGRGVAGPPGAVGLPGKDGACAGPQDGRGPPGAPGAGERGEGQPGAG 623
 QY 170 SFGQGLPGAPGPPGAGKPGEGQGVPGDLGAPGSGPAG 208
 Db 624 GPGFQGLPGSPGAGESGKPGEGQGVPGDV---GPGSGPAG 659
 RESULT 9
 Q28396
 ID Q28396 PRELIMINARY; PRT; 1418 AA.
 AC Q28396;
 DT 01-NOV-1996 (TrEMBLrel. 01, Created)
 DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
 DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
 DE Type II collagen.
 OS Equus caballus (Horse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Perissodactyla; Equidae; Equus.
 OX NCBI_TaxID=9796;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Richardson D.W.; Dodge G.R.;
 RA Submitted (JUN-1996) to the EMBL/GenBank/DBJ databases.
 RN [2]
 RP SEQUENCE OF 18-68 FROM N.A.
 RA MacLeod J.N.; Fubini S.L.; Gu D.N.; Tetraault J.W.; Todhunter R.J.;
 RL Submitted (DEC-1997) to the EMBL/GenBank/DBJ databases.
 DR EMBL; U62528; AAB05773.1; -.
 DR EMBL; AF040638; AAB96768.1; -.
 DR PIR; T45467; T45467.
 DR GO; GO:0005581; C:collagen; IEA.
 DR GO; GO:0005201; F:extracellular matrix structural constituent; IEA.
 DR InterPro; IPR008161; Clg helix.
 DR InterPro; IPR008160; Collagen.
 DR InterPro; IPR000885; Fib_collagen_C.
 DR Pfam; PF01410; COLFI; 1.
 DR Pfam; PF01391; Collagen; 18.
 DR ProDom; PD000007; Clg helix; 4.
 DR ProDom; PD002078; Fib_collagen_C; 1.
 DR SMART; SM00038; COLFI; 1.
 KW Collagen.

SQ SEQUENCE 1418 AA; 134343 MW; 115FCD19EB8696A3 CRC64;
 Query Match 72.6%; Score 860; DB 6; Length 1418;
 Best Local Similarity 74.0%; Pred. No. 7.9e-52;
 Matches 154; Conservative 10; Mismatches 44; Indels 0; Gaps 0;
 QY 1 GPFGEPTGLPGPPGGERGSGRFGPGADGVAGPKGAPGGERGSGPGAGRP 60
 Db 423 GARGEPGAGPVGPPGGERGAPGNRGGQDGLAGPKGAPGERGSGLAGPKGANGDPGRP 482
 QY 61 GEAGLPKAKLTGSPGSGPDGKTGPPGAGQDGRGPPGPGARGQAGVMGFFGPKGAA 120
 Db 483 GEPGLFQARGLTGRPDAGPQGVGSGAPGDEGREGPPGPPGQARGQPPGMVGFPGPKGAN 542
 QY 121 GEPGKAGERGVPGPGAVGAGPAGKDGSEAGAQGPPGAPGAGERGEGQPGAGSPQFQGLPGPA 180
 Db 543 GEPKAGEXGLPGAPGLRGLPGKDGTTGAAGPPGAPGAGERGEGQAGPSPGQFGLPGPP 602
 QY 181 GPPGEAGKPGEGQVPGDLGAPGSPGAG 208
 Db 603 GPPGEGKPGDQGVPGEGAGAPGLVGRG 630
 RESULT 10
 O77753
 ID O77753 PRELIMINARY; PRT; 1487 AA.
 AC O77753;
 DT 01-NOV-1998 (TrEMBLrel. 08, Created)
 DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
 DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
 DE Type IIA procollagen.
 OS Canis familiaris (Dog).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Carnivora; Fissipedia; Canidae; Canis.
 OX NCBI_TaxID=9615;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA MEDLINE=98340920; PubMed=9676231;
 RA Du F.; Acland G.M.; Ray J.;
 RT "Differential splicing of type II procollagen mRNA in canine retina."
 RL Anim. Biotechnol. 9:15-20 (1998).
 RN [2]
 RP SEQUENCE FROM N.A.
 RA MEDLINE=20480698; PubMed=11024291;
 RA Du F.; Acland G.M.; Ray J.;
 RT "Cloning and expression of type II collagen mRNA: evaluation as a
 RT candidate for canine oculo-skeletal dysplasia."
 RL Gene 255:307-316 (2000).
 DR EMBL; AF023169; AAC62178.2; -.
 DR GO; GO:0005581; C:collagen; IEA.
 DR GO; GO:0005201; F:extracellular matrix structural constituent; IEA.
 DR InterPro; IPR008161; Clg helix.
 DR InterPro; IPR008160; Collagen.
 DR InterPro; IPR000885; Fib_collagen_C.
 DR InterPro; IPR001007; VWF_C.
 DR Pfam; PF01410; COLFI; 1.
 DR Pfam; PF01391; Collagen; 18.
 DR Pfam; PF00093; vwc; 1.
 DR ProDom; PD000007; Clg helix; 4.
 DR ProDom; PD002078; Fib_collagen_C; 1.
 DR SMART; SM00038; COLFI; 1.
 DR SMART; SM00214; VWC; 1.
 DR PROSITE; PS01208; VWF_C_1; 1.
 DR PROSITE; PS01184; VWF_C_2; 1.
 KW Collagen.
 SQ SEQUENCE 1487 AA; 141875 MW; 25873EAC1E311DB8 CRC64;
 Query Match 72.8%; Score 860; DB 6; Length 1487;
 Best Local Similarity 74.0%; Pred. No. 8.2e-52;
 Matches 154; Conservative 10; Mismatches 44; Indels 0; Gaps 0;
 QY 1 GPFGEPTGLPGPPGGERGSGRFGPGADGVAGPKGAPGGERGSGPGAGRP 60

FT SIGNAL 1 25 POTENTIAL.
SQ SEQUENCE 1420 AA; 134999 MW; 88D9AAB17F214FF5 CRC64;

Query Match 72.6%; Score 859; DB 13; Length 1420;
Best Local Similarity 73.6%; Pred. No. 9.3e-52; Indels 0; Gaps 0;
Matches 153; Conservative 11; Mismatches 44;

QY 1 GPPGEPGTGLPGPPGGRGSGRFPAGDGVAGPKGAPGAGRGSPGAGPKGSPGEAGRP 60
Db 425 GARGEPGAAGVPVPPGGERGAPNGRFPQDGLAGPKGAPGERGPAGLAGPKGATGDPGRP 484

QY 61 GEAGLPGAKGLTSGSPGPGKTPGCPAGODGRCPPGPPGARGQAGVMPGPKGAA 120
Db 495 GEPGLPGARGUTGARGDAGPPGKTPGAPEDGRPPGPGQARGQPGVMPGPPGPKGAN 544

QY 121 GEPKAGRGVPGPPGAVGPPAGKDGAGAGQPPGAPGAGRGQPGAGSPFGQGLPGPA 180
Db 545 GEPKAGRGVPGPPGAVGPPAGKDGAGAGQPPGAPGAGRGQPGAGSPFGQGLPGPP 604

QY 181 GPPGKAGRGVPGPPGAVGPPAGKDGAGAGQPPGAPGAGRGQPGAGSPFGQGLPGPA 208
Db 605 GPPGKAGRGVPGPPGAVGPPAGKDGAGAGQPPGAPGAGRGQPGAGSPFGQGLPGPP 632

RESULT 14
Q14047 PRELIMINARY; PRT; 1487 AA.

AC Q14047;
DT 01-NOV-1996 (TRENBLrel. 01, Created)
DT 01-NOV-1996 (TRENBLrel. 01, Last sequence update)
DT 01-JUN-2003 (TRENBLrel. 24, Last annotation update)
DE Alpha-1 type II collagen.
CN COL2A1.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Blood;
RX MEDLINE=8519534; PubMed=3857598;
RA Cheah K.S., Stoker N.G., Griffin J.R., Grosveld F.G., Solomon E.;
RT "Identification and characterization of the human type II collagen
RT gene (COL2A1).";
RL Proc. Natl. Acad. Sci. U.S.A. 82:2555-2559(1985).
RN [2]

RP SEQUENCE FROM N.A.
RC TISSUE=Blood;
RX MEDLINE=90026318; PubMed=2803268;
RA Baldwin C.T., Reginato A.M., Smith C., Jimenez S.A., Prockop D.J.;
RT "Structure of cDNA clones coding for human type II procollagen. The
RT alpha 1(II) chain is more similar to the alpha 1(I) chain than two
RT other alpha chains of fibrillar collagens.";
RL Biochem. J. 262:521-528(1989).
RN [3]

RP SEQUENCE FROM N.A.
RC TISSUE=Blood;
RX MEDLINE=89325561; PubMed=2753125;
RA Vikkula M., Peltonen L.;
RT "Structural analyses of the polymorphic area in type II collagen
RT gene.";
RL FEBS Lett. 250:171-174(1989).
RN [4]

RP SEQUENCE FROM N.A.
RC TISSUE=Blood;
RX MEDLINE=91184811; PubMed=2081599;
RA Ryan M.C., Sieracki M., Sandell L.J.;
RT "The human type II procollagen gene: identification of an additional
RT protein-coding domain and location of potential regulatory sequences
RT in the promoter and first intron.";
RL Genomics 8:41-48(1990).
RN [5]

RP SEQUENCE FROM N.A.

RC TISSUE=Blood;
RX MEDLINE=91153296; PubMed=1999183;
RA Huang M.C., Seyer J.M., Thompson J.P., Spinella D.G., Cheah K.S.,
RA Kang A.H.;
RT "Genomic organization of the human procollagen alpha 1(II) collagen
RT gene.";
RL Eur. J. Biochem. 195:593-600(1991).
RN [6]

RP SEQUENCE FROM N.A.
RC TISSUE=Blood;
RX MEDLINE=92344585;
RA Vikkula M., Metsaranta M., Syvanen A.C., Ala-Kokko L., Vuorio E.,
RA Peltonen L.;
RT "Structural analysis of the regulatory elements of the type-II
RT procollagen gene. Conservation of promoter and first intron sequences
RT between human and mouse.";
RL Biochem. J. 285:0-0(0).
RN [7]

RP SEQUENCE FROM N.A.
RC TISSUE=Blood;
RX MEDLINE=97104294; PubMed=8948452;
RA Ala-Kokko L., Kvist A.P., Metsaranta M., Kivirikko K.I.,
RA de Crombrughe B., Prockop D.J., Vuorio E.;
RT "Conservation of the sizes of 53 introns and over 100 intronic
RT sequences for the binding of common transcription factors in the human
RT and mouse genes for type II procollagen (COL2A1).";
RL Biochem. J. 308:0-0(0).
DR ENBL; L10347; AAC41772.1; -;
DR GO; GO:0005581; C:collagen; IEA.
DR GO; GO:0005201; F:extracellular matrix structural constituent; IEA.
DR InterPro; IPR008161; Clg helix.
DR InterPro; IPR008160; Collagen.
DR InterPro; IPR000850; Fib collagen_C.
DR InterPro; IPR001007; VWF_C.
DR Pfam; PF01410; COLFI; 1.
DR Pfam; PF01391; Collagen; 18.
DR Pfam; PF00093; vwc; 1.
DR ProDom; PD000007; Clg helix; 4.
DR ProDom; PD002078; Fib collagen_C; 1.
DR SMART; SM00038; COLFI; 1.
DR SMART; SM00214; VWC; 1.
DR PROSITE; PS01208; WFC_1; 1.
DR PROSITE; PS0184; WFC_2; 1.
KW Collagen.
SQ SEQUENCE 1487 AA; 141772 MW; 0B7E79D46BDAFA97 CRC64;

Query Match 72.6%; Score 859; DB 4; Length 1487;
Best Local Similarity 74.0%; Pred. No. 9.6e-52;
Matches 154; Conservative 10; Mismatches 44; Indels 0; Gaps 0;

QY 1 GPPGEPGTGLPGPPGGRGSGRFPAGDGVAGPKGAPGAGRGSPGAGPKGSPGEAGRP 60
Db 492 GARGEPGAAGVPVPPGGERGAPNGRFPQDGLAGPKGAPGERGPAGLAGPKGANGDPGRP 551

QY 61 GEAGLPGAKGLTSGSPGPGKTPGCPAGODGRCPPGPPGARGQAGVMPGPKGAA 120
Db 552 GEPGLPGARGUTGARGDAGPPGKTPGAPEDGRPPGPGQARGQPGVMPGPPGPKGAN 611

QY 121 GEPKAGRGVPGPPGAVGPPAGKDGAGAGQPPGAPGAGRGQPGAGSPFGQGLPGPA 180
Db 612 GEPKAGRGVPGPPGAVGPPAGKDGAGAGQPPGAPGAGRGQPGAGSPFGQGLPGPP 671

QY 181 GPPGKAGRGVPGPPGAVGPPAGKDGAGAGQPPGAPGAGRGQPGAGSPFGQGLPGPA 208
Db 672 GPPGKAGRGVPGPPGAVGPPAGKDGAGAGQPPGAPGAGRGQPGAGSPFGQGLPGPP 699

RESULT 15
Q91717 PRELIMINARY; PRT; 1486 AA.
AC Q91717;
DT 01-NOV-1996 (TRENBLrel. 01, Created)
DT 01-NOV-1996 (TRENBLrel. 01, Last sequence update)

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: September 24, 2004, 11:07:31 ; Search time 25,8127 Seconds
(without alignments)
2603.588 Million cell updates/sec

Title: US-10-658-989a-1

Perfect score: 1184

Sequence: 1 GPPGEPPTGLPGRGG.....GEQGVGDLGAPGSPAGG 209

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1349238 seqs, 32158718 residues

Total number of hits satisfying chosen parameters: 1349238

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Published Applications AA:*

- 1: /cgn2_6/ptodata/1/pubaa/US07_PUBCOMB.pep:*
- 2: /cgn2_6/ptodata/1/pubaa/PCT_NEW_PUB.pep:*
- 3: /cgn2_6/ptodata/1/pubaa/US06_NEW_PUB.pep:*
- 4: /cgn2_6/ptodata/1/pubaa/US06_PUBCOMB.pep:*
- 5: /cgn2_6/ptodata/1/pubaa/US07_NEW_PUB.pep:*
- 6: /cgn2_6/ptodata/1/pubaa/PCTUS_PUBCOMB.pep:*
- 7: /cgn2_6/ptodata/1/pubaa/US08_NEW_PUB.pep:*
- 8: /cgn2_6/ptodata/1/pubaa/US08_PUBCOMB.pep:*
- 9: /cgn2_6/ptodata/1/pubaa/US09A_PUBCOMB.pep:*
- 10: /cgn2_6/ptodata/1/pubaa/US09B_PUBCOMB.pep:*
- 11: /cgn2_6/ptodata/1/pubaa/US09C_PUBCOMB.pep:*
- 12: /cgn2_6/ptodata/1/pubaa/US09C_NEW_PUB.pep:*
- 13: /cgn2_6/ptodata/1/pubaa/US10A_PUBCOMB.pep:*
- 14: /cgn2_6/ptodata/1/pubaa/US10B_PUBCOMB.pep:*
- 15: /cgn2_6/ptodata/1/pubaa/US10C_PUBCOMB.pep:*
- 16: /cgn2_6/ptodata/1/pubaa/US10C_NEW_PUB.pep:*
- 17: /cgn2_6/ptodata/1/pubaa/US60_NEW_PUB.pep:*
- 18: /cgn2_6/ptodata/1/pubaa/US60_PUBCOMB.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1149	97.0	501	12	US-10-232-175-17
2	1149	97.0	1057	16	US-10-104-889-16
3	1149	97.0	1057	16	US-10-104-889-20
4	1149	97.0	1107	16	US-10-104-889-11
5	1149	97.0	1169	16	US-10-104-889-6
6	1149	97.0	1171	16	US-10-104-889-8
7	1149	97.0	1341	14	US-10-058-224-18
8	1149	97.0	1388	16	US-10-104-889-10
9	1149	97.0	1461	16	US-10-468-091-25
10	1149	97.0	1464	12	US-09-918-715-261
11	1149	97.0	1464	14	US-10-060-036-159
12	1149	97.0	1464	14	US-10-171-311-36
13	1149	97.0	1464	14	US-10-216-705-21
14	1149	97.0	1464	14	US-10-149-352-2
15	1149	97.0	1464	14	US-10-177-293-65

16	1149	97.0	1464	14	US-10-301-822-28	Sequence 28, Appl
17	1149	97.0	1464	15	US-10-291-265-243	Sequence 243, Appl
18	1149	97.0	1464	16	US-10-357-851-1	Sequence 1, Appl
19	1149	97.0	1464	16	US-10-358-024-1	Sequence 1, Appl
20	1149	97.0	1464	16	US-10-734-564-79	Sequence 79, Appl
21	1144	96.6	1449	15	US-10-402-089-8	Sequence 8, Appl
22	1144	96.6	1449	15	US-10-402-072A-8	Sequence 8, Appl
23	1141	96.4	1463	15	US-10-402-089-2	Sequence 2, Appl
24	1141	96.4	1463	15	US-10-402-072A-2	Sequence 2, Appl
25	1116	94.3	595	15	US-10-342-331-48	Sequence 48, Appl
26	1116	94.3	595	15	US-10-342-331-50	Sequence 50, Appl
27	1116	94.3	822	15	US-10-342-331-49	Sequence 49, Appl
28	1116	94.3	822	15	US-10-468-091-26	Sequence 26, Appl
29	908	76.7	492	16	US-10-639-286-12	Sequence 12, Appl
30	866	73.1	1017	16	US-10-639-286-10	Sequence 10, Appl
31	862	72.8	1418	14	US-10-058-124-20	Sequence 20, Appl
32	862	72.8	1418	16	US-10-468-091-5	Sequence 5, Appl
33	851	71.9	1014	14	US-10-194-441A-1	Sequence 1, Appl
34	841.5	71.1	714	14	US-10-233-885-44	Sequence 44, Appl
35	841.5	71.1	714	14	US-10-231-581-44	Sequence 44, Appl
36	841.5	71.1	714	15	US-10-326-508A-15	Sequence 15, Appl
37	839	70.9	1014	14	US-10-194-441A-48	Sequence 48, Appl
38	839	70.9	1487	16	US-10-468-091-6	Sequence 6, Appl
39	814	68.8	185	12	US-10-232-175-20	Sequence 20, Appl
40	814	68.8	251	12	US-10-232-175-21	Sequence 21, Appl
41	814	68.8	500	12	US-10-232-175-22	Sequence 22, Appl
42	814	68.8	662	12	US-10-232-175-33	Sequence 33, Appl
43	787.5	66.5	1466	15	US-10-402-089-12	Sequence 12, Appl
44	787.5	66.5	1466	15	US-10-402-072A-12	Sequence 12, Appl
45	786	66.4	1496	14	US-10-177-293-70	Sequence 70, Appl

ALIGNMENTS

RESULT 1
US-10-232-175-17
; Sequence 17, Application US/10232175
; Publication No. US2003064074A1
; GENERAL INFORMATION:
; APPLICANT: Chang, Robert C.
; APPLICANT: Kivirikko, Kari I.
; APPLICANT: Neff, Thomas B.
; APPLICANT: Olsen, David R.
; APPLICANT: Polarek, James W.
; TITLE OF INVENTION: RECOMBINANT GELATINS IN VACCINES
; FILE REFERENCE: FG0224
; CURRENT APPLICATION NUMBER: US/10/232,175
; PRIOR FILING DATE: 2002-08-30
; CURRENT FILING DATE: 2002-08-30
; PRIOR FILING DATE: 2000-11-10
; NUMBER OF SEQ ID NOS: 33
; SOFTWARE: Patent in Ver. 2.0
; SEQ ID NO 17
; LENGTH: 501
; TYPE: PRT
; ORGANISM: human
US-10-232-175-17

Query Match	97.0%	Score 1149;	DB 12;	Length 501;
Best Local Similarity	93.6%	Pred. No. 7.5e-59;		
Matches 206;	Conservative	0;	Mismatches 2;	Indels 12;
Gaps	1;			
Qy	1	GPP-----GEPGPTGLPGRPGRGGPGRGFFGADGVAGPKGPKAGERSGPGA 48		
Db	280	GPPGAGEGKRGARSEPFTGLPGRPGRGGPGRGFFGADGVAGPKGPKAGERSGPGA 339		
Qy	49	GPXGSGEAGRPGEAGLPGAKGLTSGPSGPDGKGTGPPGACQDGRPPGPPGARGOA 108		
Db	340	GPXGSGEAGRPGEAGLPGAKGLTSGPSGPDGKGTGPPGACQDGRPPGPPGARGOA 399		
Qy	109	GVWGFPKGAAGEPKKAGERGVPPGAVGAGKDGAGAGCPGPGAGPAGRGFGGPA 168		

Db 400 GWMGPPGKGAAGEPCKAGRGVPGPPGAVGPAKDGAGAGQGGPAGPAGERGEGCPA 459

QY 169 GSPGQGLPGPAGPGEAGKPGEGVPGDGLGAPGSPGAG 208
 Db 460 GSPGQGLPGPAGPGEAGKPGEGVPGDGLGAPGSPGAG 499

RESULT 2

US-10-104-889-16
 ; Sequence 16, Application US/10104889
 ; Publication No. US20040086961A1
 ; GENERAL INFORMATION:

APPLICANT: GRUSKIN, ELLIOT A.
 BROKAW, JANE
 ZHANG, GUANGHUI
 PAOLELLA, DAVID

TITLE OF INVENTION: AMINO ACID MODIFIED POLYPEPTIDES
 NUMBER OF SEQUENCES: 50
 CORRESPONDENCE ADDRESS:

ADDRESSEE: DILWORTH & BARRESE
 STREET: 333 EARLE OVINGTON BOULEVARD
 CITY: UNIONDALE
 STATE: NY
 COUNTRY: U.S.A.
 ZIP: 11553

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk
 COMPUTER: IBM PC compatible
 OPERATING SYSTEM: PC-DOS/MS-DOS
 SOFTWARE: Patent in Release #1.0, Version #1.30
 CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/10/104,889
 FILING DATE: 22-Mar-2002
 CLASSIFICATION: <unknown>

PRIOR APPLICATION DATA:
 APPLICATION NUMBER: US/09/169,768
 FILING DATE: 09-OCT-1998

ATTORNEY/AGENT INFORMATION:

NAME: STEEN, JEFFREY S
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: (516) 228-8484
 TELEFAX: (516) 228-8516

INFORMATION FOR SEQ ID NO: 16:

SEQUENCE CHARACTERISTICS:
 LENGTH: 1057 amino acids
 TYPE: amino acid
 STRANDEDNESS: single
 TOPOLOGY: unknown

MOLECULE TYPE: peptide

SEQUENCE DESCRIPTION: SEQ ID NO: 16:

US-10-104-889-16

Query Match 97.0%; Score 1149; DB 16; Length 1057;
 Best Local Similarity 93.6%; Pred. No. 1.4e-58;
 Matches 206; Conservative 0; Mismatches 2; Indels 12; Gaps 1;

QY 1 GPP-----GEPGPTGLPGPPGGRGPGSRGPPGADGVAGPKGPAGERGSPGPA 48
 Db 297 GPPGAGEGKRGAGEGPGTGLPGPPGGRGPGSRGPPGADGVAGPKGPAGERGSPGPA 356

QY 49 GPKSGPAGRPGEAGLPKAGLGTGSPGSPGPDGKTGPPGAGDGRPPGPPGARGQA 108
 Db 357 GPKSGPAGRPGEAGLPKAGLGTGSPGSPGPDGKTGPPGAGDGRPPGPPGARGQA 416

QY 109 GVMGPPGPKGAAGBPGKAGRGVPGPPGAVGPAKDGEGAGAQGGPAGPAGERGSGCPA 168
 Db 417 GVMGPPGPKGAAGBPGKAGRGVPGPPGAVGPAKDGEGAGAQGGPAGPAGERGSGCPA 476

QY 169 GSPGQGLPGPAGPGEAGKPGEGVPGDGLGAPGSPGAG 208
 Db 477 GSPGQGLPGPAGPGEAGKPGEGVPGDGLGAPGSPGAG 516

RESULT 3

US-10-104-889-20
 ; Sequence 20, Application US/10104889
 ; Publication No. US20040086961A1
 ; GENERAL INFORMATION:

APPLICANT: GRUSKIN, ELLIOT A.
 BROKAW, JANE
 ZHANG, GUANGHUI
 PAOLELLA, DAVID

TITLE OF INVENTION: AMINO ACID MODIFIED POLYPEPTIDES
 NUMBER OF SEQUENCES: 50
 CORRESPONDENCE ADDRESS:

ADDRESSEE: DILWORTH & BARRESE
 STREET: 333 EARLE OVINGTON BOULEVARD
 CITY: UNIONDALE
 STATE: NY
 COUNTRY: U.S.A.
 ZIP: 11553

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk
 COMPUTER: IBM PC compatible
 OPERATING SYSTEM: PC-DOS/MS-DOS
 SOFTWARE: Patent in Release #1.0, Version #1.30
 CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/10/104,889
 FILING DATE: 22-Mar-2002
 CLASSIFICATION: <unknown>

PRIOR APPLICATION DATA:
 APPLICATION NUMBER: US/09/169,768
 FILING DATE: 09-OCT-1998

ATTORNEY/AGENT INFORMATION:

NAME: STEEN, JEFFREY S
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: (516) 228-8484
 TELEFAX: (516) 228-8516

INFORMATION FOR SEQ ID NO: 20:

SEQUENCE CHARACTERISTICS:
 LENGTH: 1057 amino acids
 TYPE: amino acid
 STRANDEDNESS: single
 TOPOLOGY: unknown

MOLECULE TYPE: peptide

SEQUENCE DESCRIPTION: SEQ ID NO: 20:

US-10-104-889-20

Query Match 97.0%; Score 1149; DB 16; Length 1057;
 Best Local Similarity 93.6%; Pred. No. 1.4e-58;
 Matches 206; Conservative 0; Mismatches 2; Indels 12; Gaps 1;

QY 1 GPP-----GEPGPTGLPGPPGGRGPGSRGPPGADGVAGPKGPAGERGSPGPA 48
 Db 297 GPPGAGEGKRGAGEGPGTGLPGPPGGRGPGSRGPPGADGVAGPKGPAGERGSPGPA 356

QY 49 GPKSGPAGRPGEAGLPKAGLGTGSPGSPGPDGKTGPPGAGDGRPPGPPGARGQA 108
 Db 357 GPKSGPAGRPGEAGLPKAGLGTGSPGSPGPDGKTGPPGAGDGRPPGPPGARGQA 416

QY 109 GVMGPPGPKGAAGBPGKAGRGVPGPPGAVGPAKDGEGAGAQGGPAGPAGERGSGCPA 168
 Db 417 GVMGPPGPKGAAGBPGKAGRGVPGPPGAVGPAKDGEGAGAQGGPAGPAGERGSGCPA 476

QY 169 GSPGQGLPGPAGPGEAGKPGEGVPGDGLGAPGSPGAG 208
 Db 477 GSPGQGLPGPAGPGEAGKPGEGVPGDGLGAPGSPGAG 516

RESULT 4

US-10-104-889-11
 ; Sequence 11, Application US/10104889
 ; Publication No. US20040086961A1
 ; GENERAL INFORMATION:

GENERAL INFORMATION:
APPLICANT: GRUSKIN, ELLIOT A.
BUECHTER, DOUGLAS
BROKAW, JANE
ZHANG, GUANGHUI
PAOLELLA, DAVID
TITLE OF INVENTION: AMINO ACID MODIFIED POLYPEPTIDES
NUMBER OF SEQUENCES: 50
CORRESPONDENCE ADDRESS:
ADDRESSEE: DILWORTH & BARRESE
STREET: 333 EARLE OVINGTON BOULEVARD
CITY: UNIONDALE

```

; COUNTRY: U.S.A.
; ZIP: 11553
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/10/104,889
; FILING DATE: 22-Mar-2002
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/09/169,768
; FILING DATE: 09-OCT-1998
; ATTORNEY/AGENT INFORMATION:
; NAME: STEEN, JEFFREY S
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (516) 228-8484
; TELEFAX: (516) 228-8516
; INFORMATION FOR SEQ ID NO: 8:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1171 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: unknown
; MOLECULE TYPE: peptide
; SEQUENCE DESCRIPTION: SEQ ID NO: 8:
US-10-104-889-8

Query Match 97.0%; Score 1149; DB 16; Length 1171;
Best Local Similarity 93.6%; Pred. No. 1.5e-56;
Matches 206; Conservative 0; Mismatches 2; Indels 12; Gaps 1;

QY 1 GPP-----GEPGTGLPCCGGERGSGRFPFGADGVAGPKGPGAGERSGPGA 48
Db 297 GPPGAGEGKRGARGEPGPTGLPCCGGERGSGRFPFGADGVAGPKGPGAGERSGPGA 356
QY 49 GPKSGPGEAGRGEAGLPCAKGLTSGSGSPGPDGKTGPPGAGQDGRPPGPPGARGQA 108
Db 357 GPKSGPGEAGRGEAGLPCAKGLTSGSGSPGPDGKTGPPGAGQDGRPPGPPGARGQA 416
QY 109 GVMFPFGKGAAGPKGAGRGVGPAGVAGPAGKDGAGAGQGGPPGAGPAGERGEGGPA 168
Db 417 GVMFPFGKGAAGPKGAGRGVGPAGVAGPAGKDGAGAGQGGPPGAGPAGERGEGGPA 476
QY 169 GSPFGQGLPGPAGPPGAGKPGEGQGVPGDLGAPGSPGAG 208
Db 477 GSPFGQGLPGPAGPPGAGKPGEGQGVPGDLGAPGSPGARG 516

RESULT 7
US-10-058-124-18
; Sequence 18, Application US/10058124
; Publication No. US20030119058A1
; GENERAL INFORMATION:
; APPLICANT: Qvist, Per
; Bonde, Martin
; TITLE OF INVENTION: A Method for Assaying Collagen Fragments
; in Body Fluids, A Test Kit and Means for Carrying Out the
; Method and Use of the Method to Diagnose the Presence of
; Disorders Associated with the Metabolism of
;
; NUMBER OF SEQUENCES: 21
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Darby & Darby PC
; STREET: 805 Third Avenue
; CITY: New York
; STATE: New York
; COUNTRY: USA
; ZIP: 10022
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS

```

```

; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/10/058,124
; FILING DATE: 29-Jan-2002
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 09/570,573
; FILING DATE: 2002-MAY-12
; APPLICATION NUMBER: 08/187,319
; FILING DATE: <Unknown>
; ATTORNEY/AGENT INFORMATION:
; NAME: Gogoris, Adda C
; REGISTRATION NUMBER: 29,714
; REFERENCE/DOCKET NUMBER: 4305/08701
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 212-527-7700
; TELEFAX: 212-753-6237
; TELEX: 236687
; INFORMATION FOR SEQ ID NO: 18:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1341 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; ORIGINAL SOURCE:
; ORGANISM: Homo sapiens
; IMMEDIATE SOURCE:
; CLONE: COLLAGEN ALPHA 1 (I)
; SEQUENCE DESCRIPTION: SEQ ID NO: 18:
US-10-058-124-18

Query Match 97.0%; Score 1149; DB 14; Length 1341;
Best Local Similarity 93.6%; Pred. No. 1.7e-58;
Matches 206; Conservative 0; Mismatches 2; Indels 12; Gaps 1;

QY 1 GPP-----GEPGTGLPCCGGERGSGRFPFGADGVAGPKGPGAGERSGPGA 48
Db 334 GPPGAGEGKRGARGEPGPTGLPCCGGERGSGRFPFGADGVAGPKGPGAGERSGPGA 393
QY 49 GPKSGPGEAGRGEAGLPCAKGLTSGSGSPGPDGKTGPPGAGQDGRPPGPPGARGQA 108
Db 394 GPKSGPGEAGRGEAGLPCAKGLTSGSGSPGPDGKTGPPGAGQDGRPPGPPGARGQA 453
QY 109 GVMFPFGKGAAGPKGAGRGVGPAGVAGPAGKDGAGAGQGGPPGAGPAGERGEGGPA 168
Db 454 GVMFPFGKGAAGPKGAGRGVGPAGVAGPAGKDGAGAGQGGPPGAGPAGERGEGGPA 513
QY 169 GSPFGQGLPGPAGPPGAGKPGEGQGVPGDLGAPGSPGAG 208
Db 514 GSPFGQGLPGPAGPPGAGKPGEGQGVPGDLGAPGSPGARG 553

RESULT 8
US-10-104-889-10
; Sequence 10, Application US/10104889
; Publication No. US20040086961A1
; GENERAL INFORMATION:
; APPLICANT: GRUSKIN, ELLIOT A.
; BUECHTER, DOUGLAS
; BROKAW, JANE
; ZHANG, GUANGHUI
; PAOLELLA, DAVID
; TITLE OF INVENTION: AMINO ACID MODIFIED POLYPEPTIDES
; NUMBER OF SEQUENCES: 50
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: DILLWORTH & BARRESE
; STREET: 333 EARLE OVINGTON BOULEVARD
; CITY: NY
; STATE: NY
; COUNTRY: U.S.A.
; ZIP: 11553
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk

```


DB 455 GPPCPAGEBCKRGARCEPPTGLPGPPGRRGGSGRFFGADGAVGPKGPAGERGSGPGA 514
 QY 49 GPKGSPGEAGRCEAGLPGAKGLTGSFGSPGPDGKTGPPGPAGQDGRPPGPPGARGQA 108
 DB 515 GPKGSPGEAGRCEAGLPGAKGLTGSFGSPGPDGKTGPPGPAGQDGRPPGPPGARGQA 574
 QY 109 GVMGRRPGPKGAAGEPKKAGRGVPGPPGAVGPGAKDGEAGAQAQPPGPPAGPAGERGEGGPA 168
 DB 575 GVMGRRPGPKGAAGEPKKAGRGVPGPPGAVGPGPKDGEAGAQAQPPGPPAGPAGERGEGGPA 634
 QY 169 GSPGFQCLPGPAGPPGAEAGKPGSQGVPGDLGAPGPSGPA 208
 DB 635 GSPGFQCLPGPAGPPGAEAGKPGSQGVPGDLGAPGPSGARG 674
 RESULT 10
 US-09-918-715-261
 ; Sequence 261, Application US/09918715
 ; Publication No. US20030017157A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Brad St. Croix
 ; APPLICANT: Bert Vogelstein
 ; APPLICANT: Kenneth Kinzler
 ; TITLE OF INVENTION: ENDOTHELIAL CELL EXPRESSION PATTERNS
 ; FILE REFERENCE: 1107.00134
 ; CURRENT APPLICATION NUMBER: US/09/918,715
 ; CURRENT FILING DATE: 2001-08-01
 ; PRIOR APPLICATION NUMBER: 60/222,599
 ; PRIOR FILING DATE: 2000-08-02
 ; PRIOR APPLICATION NUMBER: 60/224,360
 ; PRIOR FILING DATE: 2000-08-11
 ; PRIOR APPLICATION NUMBER: 60/282,950
 ; PRIOR FILING DATE: 2000-04-11
 ; NUMBER OF SEQ ID NOS: 358
 ; SOFTWARE: Fast-Seq for Windows Version 3.0
 ; SEQ ID NO 261
 ; LENGTH: 1464
 ; TYPE: PRT
 ; ORGANISM: Homo sapiens
 US-09-918-715-261
 Query Match 97.0%; Score 1149; DB 12; Length 1464;
 Best Local Similarity 93.6%; Pred. No. 1.8e-58;
 Matches 206; Conservative 0; Mismatches 2; Indels 12; Gaps 1;
 QY 1 GPP-----GEPQPTGLPGPPGRRGGSGRFFGADGAVGPKGPAGERGSGPGA 48
 DB 458 GPPCPAGEBCKRGARCEPPTGLPGPPGRRGGSGRFFGADGAVGPKGPAGERGSGPGA 517
 QY 49 GPKGSPGEAGRCEAGLPGAKGLTGSFGSPGPDGKTGPPGPAGQDGRPPGPPGARGQA 108
 DB 518 GPKGSPGEAGRCEAGLPGAKGLTGSFGSPGPDGKTGPPGPAGQDGRPPGPPGARGQA 577
 QY 109 GVMGRRPGPKGAAGEPKKAGRGVPGPPGAVGPGAKDGEAGAQAQPPGPPAGPAGERGEGGPA 168
 DB 578 GVMGRRPGPKGAAGEPKKAGRGVPGPPGAVGPGPKDGEAGAQAQPPGPPAGPAGERGEGGPA 637
 QY 169 GSPGFQCLPGPAGPPGAEAGKPGSQGVPGDLGAPGPSGPA 208
 DB 638 GSPGFQCLPGPAGPPGAEAGKPGSQGVPGDLGAPGPSGARG 677
 RESULT 11
 US-10-060-036-159
 ; Sequence 159, Application US/10060036
 ; Publication No. US20030073144A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Benson, Darin R.
 ; APPLICANT: Kalos, Michael D.
 ; APPLICANT: Lodes, Michael J.
 ; APPLICANT: Persing, David H.
 ; APPLICANT: Hepler, William I.
 ; APPLICANT: Jiang, Yuqin

; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY
; FILE REFERENCE: 210121.566
; CURRENT APPLICATION NUMBER: US/10/060,036
; NUMBER OF SEQ ID NOS: 4560
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 159
; LENGTH: 1464
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-060-036-159

Query Match 97.0%; Score 1149; DB 14; Length 1464;
Best Local Similarity 93.6%; Pred. No. 1.8e-58;
Matches 206; Conservative 0; Mismatches 2; Indels 12; Gaps 1;
QY 1 GPP-----GEPGPTGLPDPGGERGSGRPPGADGVAGPKGAGERGSPGPA 48
DB 458 GPPGAGEGKRGARGEPGPTGLPDPGGERGSGRPPGADGVAGPKGAGERGSPGPA 517
QY 49 GPKSGPGEAGRPFGEAGLFGAKGLTSGPSGPDGKTGPPGAGQDGRPPGPPGARGQA 108
DB 518 GPKSGPGEAGRPFGEAGLFGAKGLTSGPSGPDGKTGPPGAGQDGRPPGPPGARGQA 577
QY 109 GVMGPPGPKGAAGBPKKAGERGVPPGAVGPKDGEAGAGQGGPPGAGPAGERGEOGPA 168
DB 578 GVMGPPGPKGAAGBPKKAGERGVPPGAVGPKDGEAGAGQGGPPGAGPAGERGEOGPA 637
QY 169 GSPGFQGLPAGPPGAGEAGKPGEGQGVPGDLGAPGSPGAG 208
DB 638 GSPGFQGLPAGPPGAGEAGKPGEGQGVPGDLGAPGSPGAG 677

RESULT 12
US-10-171-311-36
; Sequence 36, Application US/10171311
; Publication No. US20030087270A1
; GENERAL INFORMATION:
; APPLICANT: Schlegel, Robert
; APPLICANT: Chen, Yan
; APPLICANT: Zhao, Xumei
; APPLICANT: Monahan, John
; APPLICANT: Kamatkar, Shubhangi
; APPLICANT: Glatt, Karen
; APPLICANT: Gannavarapu, Manjula
; APPLICANT: Hoerish, Sebastian
; TITLE OF INVENTION: NOVEL GENES, COMPOSITIONS, KITS, AND METHODS FOR
; TITLE OF INVENTION: IDENTIFICATION, ASSESSMENT, PREVENTION, AND THERAPY
; TITLE OF INVENTION: OF CERVICAL CANCER
; FILE REFERENCE: MRI-035
; CURRENT APPLICATION NUMBER: US/10/171.311
; CURRENT FILING DATE: 2002-06-12
; PRIOR APPLICATION NUMBER: US 60/298,159
; PRIOR FILING DATE: 2001-06-13
; PRIOR APPLICATION NUMBER: US 60/298,155
; PRIOR FILING DATE: 2001-06-13
; PRIOR APPLICATION NUMBER: US 60/335,936
; PRIOR FILING DATE: 2001-11-14
; NUMBER OF SEQ ID NOS: 238
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 36
; LENGTH: 1464
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-171-311-36

Query Match 97.0%; Score 1149; DB 14; Length 1464;
Best Local Similarity 93.6%; Pred. No. 1.8e-58;
Matches 206; Conservative 0; Mismatches 2; Indels 12; Gaps 1;
QY 1 GPP-----GEPGPTGLPDPGGERGSGRPPGADGVAGPKGAGERGSPGPA 48

DB 458 GPPGAGEGKRGARGEPGPTGLPDPGGERGSGRPPGADGVAGPKGAGERGSPGPA 517
QY 49 GPKSGPGEAGRPFGEAGLFGAKGLTSGPSGPDGKTGPPGAGQDGRPPGPPGARGQA 108
DB 518 GPKSGPGEAGRPFGEAGLFGAKGLTSGPSGPDGKTGPPGAGQDGRPPGPPGARGQA 577
QY 109 GVMGPPGPKGAAGBPKKAGERGVPPGAVGPKDGEAGAGQGGPPGAGPAGERGEOGPA 168
DB 578 GVMGPPGPKGAAGBPKKAGERGVPPGAVGPKDGEAGAGQGGPPGAGPAGERGEOGPA 637
QY 169 GSPGFQGLPAGPPGAGEAGKPGEGQGVPGDLGAPGSPGAG 208
DB 638 GSPGFQGLPAGPPGAGEAGKPGEGQGVPGDLGAPGSPGAG 677

RESULT 13
US-10-216-705-21
; Sequence 21, Application US/10216705
; Publication No. US20030096973A1
; GENERAL INFORMATION:
; APPLICANT: Meristem Therapeutics, S.A.
; TITLE OF INVENTION: Recombinant Collagens and Derived Proteins Produced by Plants, Met
; FILE REFERENCE: 1149-3 DIV
; CURRENT APPLICATION NUMBER: US/10/216,705
; CURRENT FILING DATE: 2002-08-09
; PRIOR APPLICATION NUMBER: US 09/331,347
; PRIOR FILING DATE: 1999-08-17
; NUMBER OF SEQ ID NOS: 22
; SOFTWARE: Patent in version 3.1
; SEQ ID NO 21
; LENGTH: 1464
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-216-705-21

Query Match 97.0%; Score 1149; DB 14; Length 1464;
Best Local Similarity 93.6%; Pred. No. 1.8e-58;
Matches 206; Conservative 0; Mismatches 2; Indels 12; Gaps 1;
QY 1 GPP-----GEPGPTGLPDPGGERGSGRPPGADGVAGPKGAGERGSPGPA 48
DB 458 GPPGAGEGKRGARGEPGPTGLPDPGGERGSGRPPGADGVAGPKGAGERGSPGPA 517
QY 49 GPKSGPGEAGRPFGEAGLFGAKGLTSGPSGPDGKTGPPGAGQDGRPPGPPGARGQA 108
DB 518 GPKSGPGEAGRPFGEAGLFGAKGLTSGPSGPDGKTGPPGAGQDGRPPGPPGARGQA 577
QY 109 GVMGPPGPKGAAGBPKKAGERGVPPGAVGPKDGEAGAGQGGPPGAGPAGERGEOGPA 168
DB 578 GVMGPPGPKGAAGBPKKAGERGVPPGAVGPKDGEAGAGQGGPPGAGPAGERGEOGPA 637
QY 169 GSPGFQGLPAGPPGAGEAGKPGEGQGVPGDLGAPGSPGAG 208
DB 638 GSPGFQGLPAGPPGAGEAGKPGEGQGVPGDLGAPGSPGAG 677

RESULT 14
US-10-149-352-2
; Sequence 2, Application US/10149352
; Publication No. US20030105050A1
; GENERAL INFORMATION:
; APPLICANT: Beri, Rajinder
; TITLE OF INVENTION: ANTISENSE OLIGONUCLEOTIDES
; FILE REFERENCE: 06275-254US1
; CURRENT APPLICATION NUMBER: US/10/149,352
; CURRENT FILING DATE: 2002-06-10
; PRIOR APPLICATION NUMBER: PCT/GS00/04741
; PRIOR FILING DATE: 2000-12-12
; PRIOR APPLICATION NUMBER: GB 9929487.8
; PRIOR FILING DATE: 1999-12-15
; NUMBER OF SEQ ID NOS: 14
; SOFTWARE: Patent in Ver. 4.0

US-10-177-293-65

Query Match 97.0%; Score 1149; DB 14; Length 1464;
 Best Local Similarity 93.6%; Pred. No. 1.8e-58;
 Matches 206; Conservative 0; Mismatches 2; Indels 12; Gaps 1;

QY 1 GPP-----GEPGPTGLPDPGGERGGSGRGGFPAGDGVAGPKGPAGRGSPGPA 48
 DB 458 GPPGAGEGKRGARGEPGPTGLPDPGGERGGSGRGGFPAGDGVAGPKGPAGRGSPGPA 517
 QY 49 GPKSGPEAGRPGEAGLPGAKGLTGSFGSPGPDGKTGPPAGDGRPPGPPGARGQA 108
 DB 518 GPKSGPEAGRPGEAGLPGAKGLTGSFGSPGPDGKTGPPAGDGRPPGPPGARGQA 577
 QY 109 GVMGFPKGAAGFPKAGRGVFPFGCAVGPAGKDGAGAGQGGPPGAGPAGERGEGGPA 168
 DB 578 GVMGFPKGAAGFPKAGRGVFPFGCAVGPAGKDGAGAGQGGPPGAGPAGERGEGGPA 637
 QY 169 GSPGFGGLPGPAGPPGAGKDGAGAGQGGPPGAGPAGERGEGGPA 208
 DB 638 GSPGFGGLPGPAGPPGAGKDGAGAGQGGPPGAGPAGERGEGGPA 677

Search completed: September 24, 2004, 11:13:23
 Job time : 26.8127 secs

US-10-658-989a-1.rapb

Query Match 97.0%; Score 1149; DB 14; Length 1464;
 Best Local Similarity 93.6%; Pred. No. 1.8e-58;
 Matches 206; Conservative 0; Mismatches 2; Indels 12; Gaps 1;

QY 1 GPP-----GEPGPTGLPDPGGERGGSGRGGFPAGDGVAGPKGPAGRGSPGPA 48
 DB 458 GPPGAGEGKRGARGEPGPTGLPDPGGERGGSGRGGFPAGDGVAGPKGPAGRGSPGPA 517
 QY 49 GPKSGPEAGRPGEAGLPGAKGLTGSFGSPGPDGKTGPPAGDGRPPGPPGARGQA 108
 DB 518 GPKSGPEAGRPGEAGLPGAKGLTGSFGSPGPDGKTGPPAGDGRPPGPPGARGQA 577
 QY 109 GVMGFPKGAAGFPKAGRGVFPFGCAVGPAGKDGAGAGQGGPPGAGPAGERGEGGPA 168
 DB 578 GVMGFPKGAAGFPKAGRGVFPFGCAVGPAGKDGAGAGQGGPPGAGPAGERGEGGPA 637
 QY 169 GSPGFGGLPGPAGPPGAGKDGAGAGQGGPPGAGPAGERGEGGPA 208
 DB 638 GSPGFGGLPGPAGPPGAGKDGAGAGQGGPPGAGPAGERGEGGPA 677

RESULT 15

US-10-177-293-65

Sequence 65, Application US/10177293
 Publication No. US20030124128A1
 GENERAL INFORMATION:
 APPLICANT: Lillie, James
 APPLICANT: Glatt, Karen
 APPLICANT: Zhao, Xumei
 APPLICANT: Ganavaru, Manjula
 APPLICANT: Kamatkar, Shubhangi
 APPLICANT: Mertens, Maureen
 APPLICANT: Myer, Vic
 APPLICANT: Wang, Youzhen
 APPLICANT: Xu, Yongyao
 APPLICANT: Hoersch, Sebastian
 APPLICANT: Monahan, John
 APPLICANT: Meyers, Rachel E.
 APPLICANT: Bast Jr., Robert C.
 APPLICANT: Hortobagyi, Gabriel N.
 APPLICANT: Pusztai, Lajos
 APPLICANT: Meric, Funda
 APPLICANT: Sahin, Aysegul
 APPLICANT: Mills, Gordon B.
 TITLE OF INVENTION: COMPOSITIONS, KITS, AND METHODS FOR IDENTIFICATION, ASSESSMENT,
 PREVENTION, AND THERAPY OF BREAST CANCER
 FILE REFERENCE: MRI-038
 CURRENT APPLICATION NUMBER: US/10/177,293
 CURRENT FILING DATE: 2002-06-21
 PRIOR APPLICATION NUMBER: US 60/299,887
 PRIOR FILING DATE: 2001-06-21
 PRIOR APPLICATION NUMBER: US 60/301,572
 PRIOR FILING DATE: 2001-06-27
 PRIOR APPLICATION NUMBER: US 60/306,501
 PRIOR FILING DATE: 2001-07-18
 PRIOR APPLICATION NUMBER: US 60/325,002
 PRIOR FILING DATE: 2001-09-25
 PRIOR APPLICATION NUMBER: US 60/362,585
 PRIOR FILING DATE: 2002-03-05
 PRIOR APPLICATION NUMBER: US 60/xxx,xxx
 PRIOR FILING DATE: 2002-05-14
 NUMBER OF SEQ ID NOS: 506
 SOFTWARE: FastSeq for Windows Version 4.0
 SEQ ID NO 65
 LENGTH: 1464
 TYPE: PRT
 ORGANISM: Homo sapiens

This Page Blank (usp10)

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: September 24, 2004, 11:06:55 ; Search time 5.27357 Seconds
(without alignments)
2063.622 Million cell updates/sec

Title: US-10-658-989a-1

Perfect score: 1184

Sequence: 1 GPGGEPGPTGLPGPGERGG.....GEGGVFDLGAQPGSPAGG 209

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 141681 seqs, 52070155 residues

Total number of hits satisfying chosen parameters: 141681

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt_42:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1149	97.0	1460	1	Q9XSJ7 canis fami
2	1149	97.0	1464	1	P02452 homo sapien
3	1118	94.4	671	1	P02454 rattus norv
4	1116	94.3	1453	1	P1087 mus musculus
5	1067	90.1	1453	1	P02457 gallus gall
6	913	77.1	747	1	P02459 bos taurus
7	862	72.8	1418	1	P02458 homo sapien
8	839	70.9	1459	1	P28481 mus musculus
9	824	69.6	779	1	P02453 bos taurus
10	794.5	67.1	1262	1	P12105 gallus gall
11	786	66.4	1496	1	P05997 homo sapien
12	776	65.5	1466	1	P02461 homo sapien
13	764	64.5	1372	1	P02466 rattus norv
14	763	64.4	1464	1	P03121 mus musculus
15	761.5	64.3	1049	1	P04258 bos taurus
16	760	64.2	1372	1	Q01149 mus musculus
17	744	62.8	1364	1	P02465 bos taurus
18	739	62.4	1366	1	P08123 homo sapien
19	732	61.8	1366	1	O46392 canis fami
20	703	59.4	1355	1	O42350 rana catesb
21	693.5	58.6	1356	1	O93484 oncorhynch
22	650.5	54.9	1362	1	P02467 gallus gall
23	629.5	53.2	1027	1	P30754 riftia pach
24	606	51.2	1838	1	P20908 homo sapien
25	590.5	49.9	1736	1	P13942 homo sapien
26	584.5	49.4	1736	1	O64739 mus musculus
27	577	48.7	1806	1	P12107 homo sapien
28	576	48.6	754	1	Q28247 canis fami
29	576	48.6	911	1	Q28083 bos taurus
30	575.5	48.6	3124	1	P13944 gallus gall
31	574.5	48.5	675	1	P32017 gallus gall
32	573.5	48.4	1690	1	P53420 homo sapien
33	572.5	48.4	636	1	P13941 rattus norv

34	571.5	48.3	684	1	CA39_HUMAN	Q14050 homo sapien
35	571	48.2	1745	1	CA35_HUMAN	P25940 homo sapien
36	568.5	48.0	1804	1	CA1B_MOUSE	Q61245 mus musculus
37	563.5	47.6	2944	1	CA17_HUMAN	Q02388 homo sapien
38	563	47.6	1685	1	CA5A_HUMAN	P29400 homo sapien
39	561	47.4	1603	1	CA1F_HUMAN	Q07092 homo sapien
40	555.5	46.9	921	1	CA19_HUMAN	P20849 homo sapien
41	552.5	46.7	674	1	CA1A_CHICK	P08125 gallus gall
42	552.5	46.7	680	1	CA1A_HUMAN	Q03692 homo sapien
43	549	46.4	526	1	CA21_RABIT	Q28668 oryctolagus
44	548.5	46.3	921	1	CA19_MOUSE	Q05722 mus musculus
45	547	46.2	1712	1	CA24_HUMAN	P08572 homo sapien

ALIGNMENTS

RESULT 1
CALL_CANFA STANDARD; PRT; 1460 AA.
ID AC Q9XSJ7;
DT 30-MAY-2000 (Rel. 39, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 15-MAR-2004 (Rel. 43, Last annotation update)
DE Collagen alpha 1(I) chain precursor.
GN COL1A1.
OS Canis familiaris (Dog).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Carnivora; Fissipedia; Canidae; Canis.
OX NCBI_Taxid=9615;
RN [1]
RP SEQUENCE FROM N.A., AND VARIANT OI ALA-208.
RC TISSUE=Skin;
RX MEDLINE=21023337; PubMed=11147834;
RA Campbell B.G., Wootton J.A.M., MacLeod J.N., Minor R.R.;
RT Sequence of normal canine COL1A1 cDNA and identification of a
RT heterozygous alpha 1(I) collagen Gly208Ala mutation in a severe case
RT of canine osteogenesis imperfecta.;
RL Arch. Biochem. Biophys. 384:37-46(2000).
CC -I- FUNCTION: Type I collagen is a member of group I collagen
(fibrillar forming collagen).
CC -I- SUBUNIT: Trimers of one alpha 2(I) and two alpha 1(I) chains.
CC -I- PTM: Prolines at the third position of the tripeptide repeating
CC unit (G-X-Y) are hydroxylated in some or all of the chains.
CC -I- DISEASE: Defects in COL1A1 are a cause of osteogenesis imperfecta
(OI).
CC -I- SIMILARITY: Contains 1 VWFC domain.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; AF153062; AAD34619.1; -
CC InterPro; IPR008161; C1g_helix.
CC InterPro; IPR008160; Collagen.
CC InterPro; IPR000885; Fib_collagen_C.
CC InterPro; IPR002181; Fibrinogen_C.
CC InterPro; IPR001007; VWFC_C.
CC Pfam; PF01410; COLFI; 1.
CC Pfam; PF01391; Collagen; 18.
CC ProDom; PD000007; C1g_helix; 2.
CC ProDom; PD02078; Fib_collagen_C; 1.
CC SMART; SM00038; COLFI; 1.
CC SMART; SM00214; VWFC; 1.
CC PROSITE; PS01208; VWFC_1; 1.
CC PROSITE; PS01184; VWFC_2; 1.
KW Extracellular matrix; Connective tissue; Repeat; Hydroxylation;
KW Glycoprotein; Collagen; Signal; Disease mutation.
FT SIGNAL 1 22 BY SIMILARITY.

FT PROPEP 23 157 AMINO-TERMINAL PROPEPTIDE.
 FT CHAIN 158 1214 COLLAGEN ALPHA 1(I) CHAIN.
 FT PROPEP 1215 1460 CARBOXYL-TERMINAL PROPEPTIDE.
 FT DOMAIN 34 92 VWFC.
 FT DOMAIN 158 174 NONHELICAL REGION (N-TERMINAL).
 FT DOMAIN 175 1188 TRIPLE-HELICAL REGION.
 FT SITE 1189 1214 NONHELICAL REGION (C-TERMINAL).
 FT SITE 741 743 CELL ATTACHMENT SITE (POTENTIAL).
 FT SITE 1089 1091 CELL ATTACHMENT SITE (POTENTIAL).
 FT CARBOHYD 1361 1361 N-LINKED (GLCNAC...) (POTENTIAL).
 FT VARIANT 208 208 G -> A (in OI, severe).
 SQ SEQUENCE 1460 AA; 138762 MW; 58E3674D2B570697 CRC64;

Query Match 97.0%; Score 1149; DB 1; Length 1460;
 Best Local Similarity 93.6%; Pred. No. 4.1e-48;
 Matches 206; Conservative 0; Mismatches 2; Indels 12; Gaps 1;

QY 1 GPP-----GEPGPTGLPGGPGGSGRGGPGGAGVAGKPGAGRGSGGPA 48
 |||||
 Db 454 GPPGAGGEGKRGAGEGPTGLPGGPGGSGRGGPGGAGVAGKPGAGRGSGGPA 513
 |||||

QY 49 GPKSGPAGRGAGEAGLPGKAGLGTGSPGPGDGKTPPGAGQDGRPGPPGPGARGQA 108
 |||||

Db 514 GPKSGPAGRGAGEAGLPGKAGLGTGSPGPGDGKTPPGAGQDGRPGPPGARGQA 573
 |||||

QY 109 GVMGPPGKGAAGBPKAGRGVPGPGCAVCPAGKDGEGAGQPGGAGRGGEQGPA 168
 |||||

Db 574 GVMGPPGKGAAGBPKAGRGVPGPGCAVCPAGKDGEGAGQPGGAGRGGEQGPA 633
 |||||

QY 169 GSPFGQLPGGAGPPGAGKPGEGQVPGDGLGAPGSPGAG 208
 |||||

Db 634 GSPFGQLPGGAGPPGAGKPGEGQVPGDGLGAPGSPGAG 673
 |||||

RESULT 2
 CALL HUMAN
 ID CALL HUMAN STANDARD; PRT; 1464 AA.
 AC P02452; Q14037; Q15176;
 DT 21-JUL-1986 (Rel. 01, Created)
 DT 01-MAR-1989 (Rel. 10, Last sequence update)
 DT 10-OCT-2003 (Rel. 42, Last annotation update)
 DE Collagen alpha 1(I) chain precursor.
 GN COL1A1.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 OC NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE OF 1-472 FROM N.A.
 RX MEDLINE=89025644; PubMed=3178743;
 RA Tromp G., Kuivaniemi H., Stacey A., Shikata H., Baldwin C.T.,
 Jaenisch R., Prockop D.J.;
 RT "Structure of a full-length cDNA clone for the prepro alpha 1(I)
 chain of human type I procollagen.";
 RL Biochem. J. 253:919-922(1988).
 RN [2]
 RP SEQUENCE OF 1-181 FROM N.A.
 RX MEDLINE=84270697; PubMed=6462220;
 RA Chu M.-L., de Wet W.J., Bernard M.P., Ding J.-F., Morabito M.,
 Myers J., Williams C., Ramirez F.;
 RT "Human pro alpha 1(I) collagen gene structure reveals evolutionary
 conservation of a pattern of introns and exons.";
 RL Nature 310:330-340(1984).
 RN [3]
 RP SEQUENCE OF 162-301.
 TISSUE=Skin;
 RX MEDLINE=71038625; PubMed=5529814;
 RA Click E.M., Bornstein P.;
 RT "Isolation and characterization of the cyanogen bromide peptides from
 the alpha 1 and alpha 2 chains of human skin collagen.";
 RL Biochemistry 9:4699-4706(1970).
 RN [4]
 RP SEQUENCE OF 263-268.

RC TISSUE=Skin;
 RX MEDLINE=71001508; PubMed=4319110;
 RA Morgan P.H., Jacobs H.G., Segrest J.P., Cunningham L.W.;
 RT "A comparative study of glycopeptides derived from selected
 vertebrate collagens. A possible role of the carbohydrate in fibril
 formation.";
 RL J. Biol. Chem. 245:5042-5048(1970).
 RN [5]
 RP SEQUENCE OF 425-1464 FROM N.A.
 RX MEDLINE=84080385; PubMed=689127;
 RA Bernard M.P., Chu M.-L., Myers J.C., Ramirez F., Eikenberry E.F.,
 Prockop D.J.;
 RT "Nucleotide sequences of complementary deoxyribonucleic acids for the
 pro alpha 1 chain of human type I procollagen. Statistical evaluation
 of structures that are conserved during evolution.";
 RL Biochemistry 22:5213-5223(1983).
 RN [6]
 RP SEQUENCE OF 1229-1454 FROM N.A.
 TISSUE=Bone;
 RX MEDLINE=88124208; PubMed=3340531;
 RA Maekelae J.K., Raessina M., Virta A., Vuorio E.;
 RT "Human pro alpha 1(I) collagen: cDNA sequence for the C-propeptide
 domain.";
 RL Nucleic Acids Res. 16:349-349(1988).
 RN [7]
 RP SEQUENCE OF 1-34 FROM N.A.
 RX MEDLINE=88097389; PubMed=3480516;
 RA Bornstein P., McKay J., Morishima J.K., Devarayalu S., Gelinas R.E.;
 RT "Regulatory elements in the first intron contribute to
 transcriptional control of the human alpha 1(I) collagen gene.";
 RL Proc. Natl. Acad. Sci. U.S.A. 84:8869-8873(1987).
 RN [8]
 RP SEQUENCE OF 1-34 FROM N.A.
 RX MEDLINE=85130970; PubMed=2857713;
 RA Chu M.-L., de Wet W.J., Bernard M.P., Ramirez F.;
 RT "Fine structural analysis of the human pro-alpha 1(I) collagen gene.
 Promoter structure, AluI repeats, and polymorphic transcripts.";
 RL J. Biol. Chem. 260:2315-2320(1985).
 RN [9]
 RP SEQUENCE OF 1-44 FROM N.A.
 RX MEDLINE=88033098; PubMed=2822714;
 RA Rossouw C.M.S., Vergeer W.P., du Plooy S.J., Bernard M.P., Ramirez F.,
 de Wet W.J.;
 RT "DNA sequences in the first intron of the human pro-alpha 1(I)
 collagen gene enhance transcription.";
 RL J. Biol. Chem. 262:15151-15157(1987).
 RN [10]
 RP REVIEW ON VARIANTS.
 RX MEDLINE=91184577; PubMed=2010058;
 RA Kuivaniemi H., Tromp G., Prockop D.J.;
 RT "Mutations in collagen genes: causes of rare and some common diseases
 in humans.";
 RL FASEB J. 5:2052-2060(1991).
 RN [11]
 RP REVIEW ON VARIANTS.
 RX MEDLINE=91374476; PubMed=1895312;
 RA Kuivaniemi H., Tromp G., Prockop D.J.;
 RT "Mutations in fibrillar collagens (types I, II, III, and XI), fibril-
 associated collagen (type IX), and network-forming collagen (type X)
 cause a spectrum of diseases of bone, cartilage, and blood vessels.";
 RL Hum. Mutat. 9:300-315(1997).
 RN [12]
 RP REVIEW ON OI VARIANTS.
 RX MEDLINE=91374476; PubMed=1895312;
 RA Byers P.H., Wallis G.A., Willing M.C.;
 RT "Osteogenesis imperfecta: translation of mutation to phenotype.";
 RL J. Med. Genet. 28:433-442(1991).
 RN [13]
 RP REVIEW ON OI VARIANTS.
 RX MEDLINE=97169389; PubMed=9016532;
 RA Daigleish R.;
 RT "The human type I collagen mutation database.";
 RL Nucleic Acids Res. 25:181-187(1997).

[14] VARIANT OI-II CYS-1166.
RX MEDLINE=86287390; PubMed=3016737;
RA Cohn D.H., Byers P.H., Steinmann B., Gelinas R.E.;
RT "lethal osteogenesis imperfecta resulting from a single nucleotide
RL change in one human pro alpha 1(I) collagen allele.";
RL Proc. Natl. Acad. Sci. U.S.A. 83:6045-6047(1986).
[15]
RX VARIANT OI-II ARG-569.
RX MEDLINE=87222295; PubMed=3108247;
RA Bateman J.F., Chan D., Walkers I.D., Rogers J.G., Cole W.G.;
RT "lethal perinatal osteogenesis imperfecta due to the substitution of
RT arginine for glycine at residue 391 of the alpha 1(I) chain of type I
RT collagen.";
RL J. Biol. Chem. 262:7021-7027(1987).
[16]
RX VARIANT OI-II CYS-926.
RX MEDLINE=89033031; PubMed=3667599;
RA Vogel B.E., Minor R.R., Freund M., Prockop D.J.;
RT "A point mutation in a type I procollagen gene converts glycine 748
RT of the alpha 1 chain to cysteine and destabilizes the triple helix in
RT a lethal variant of osteogenesis imperfecta.";
RL J. Biol. Chem. 262:14737-14744(1987).
[17]
RX VARIANT OI-II ARG-842.
RX MEDLINE=88298828; PubMed=3403550;
RA Bateman J.F., Lamande S.R., Dahl H.H., Chan D., Cole W.G.;
RT "Substitution of arginine for glycine 664 in the collagen alpha 1(I)
RT chain in lethal perinatal osteogenesis imperfecta. Demonstration of
RT the peptide defect by in vitro expression of the mutant cDNA.";
RL J. Biol. Chem. 263:11627-11630(1988).
[18]
RX VARIANT OI CYS-1195.
RX MEDLINE=89218628; PubMed=3244312;
RA Labhard M.E., Wirtz M.K., Pope F.M., Nicholls A.C., Hollister D.W.;
RT "A cysteine for glycine substitution at position 1017 in an alpha
RT 1(I) chain of type I collagen in a patient with mild dominantly
RT inherited osteogenesis imperfecta.";
RL Mol. Biol. Med. 5:197-207(1988).
[19]
RX VARIANT OI-II VAL-434.
RX MEDLINE=89255493; PubMed=2470760;
RA Patterson E., Smiley E., Bonadio J.;
RT "RNA sequence analysis of a perinatal lethal osteogenesis imperfecta
RT mutation.";
RL J. Biol. Chem. 264:10083-10087(1989).
[20]
RX VARIANT OI-IV SER-1010.
RX MEDLINE=89308591; PubMed=2745420;
RA Marini J.C., Grange D.K., Gottesman G.S., Lewis M.B., Koepf D.A.;
RT "Osteogenesis imperfecta type IV. Detection of a point mutation in
RT one alpha 1(I) collagen allele (COL1A1) by RNA/RNA hybrid analysis.";
RL J. Biol. Chem. 264:11893-11900(1989).
[21]
RX VARIANTS OI-II ALA-1106; VAL-1151; ARG-1154 AND VAL-1184.
RX MEDLINE=89380165; PubMed=277764;
RA Lamande S.R., Dahl H.-H.M., Cole W.G., Bateman J.F.;
RT "Characterization of point mutations in the collagen COL1A1 and
RT COL1A2 genes causing lethal perinatal osteogenesis imperfecta.";
RL J. Biol. Chem. 264:15809-15812(1989).
[22]
RX VARIANT OI SER-1022.
RX MEDLINE=90062068; PubMed=25111192;
RA Pack M., Constantinou C.D., Kalia K., Nielsen K.B., Prockop D.J.;
RT "Substitution of serine for alpha 1(I)-glycine 844 in a severe
RT variant of osteogenesis imperfecta minimally destabilizes the triple
RT helix of type I procollagen. The effects of glycine substitutions on
RT thermal stability are either position of amino acid specific.";
RL J. Biol. Chem. 264:19694-19699(1989).
[23]
RX VARIANT OI-II CYS-1082.
RX MEDLINE=89109573; PubMed=2913053;
RA Constantinou C.D., Nielsen K.B., Prockop D.J.;

"A lethal variant of osteogenesis imperfecta has a single base
RT mutation that substitutes cysteine for glycine 904 of the alpha 1(I)
RT chain of type I procollagen. The asymptomatic mother has an
RT unidentified mutation producing an overmodified and unstable type I
RT procollagen.";
RL J. Clin. Invest. 83:574-584(1989).
[24]
RX VARIANTS OI CYS-272; CYS-704 AND CYS-896.
RX MEDLINE=90009313; PubMed=2794057;
RA Starman B.J., Eyre D., Charbonneau H., Harrylock M., Weis M.A.,
RA Weiss L., Graham J.M., Byers P.H.;
RT "Osteogenesis imperfecta. The position of substitution for glycine by
RT cysteine in the triple helical domain of the pro alpha 1(I) chains of
RT type I collagen determines the clinical phenotype.";
RL J. Clin. Invest. 84:1205-1214(1989).
[25]
RX VARIANT OI-II CYS-422.
Query Match 97.0%; Score 1149; DB 1; Length 1464;
Best Local Similarity 93.6%; Pred. No. 4,1e-48;
Matches 206; Conservative 0; Mismatches 2; Indels 12; Gaps 1;
QY 1 GPP-----GEPGPTCLPPDPPGSGRGGPGADGVAGPKGAGERSGPGA 48
DB 458 GPPGAGEGKRGARGEFGPTGLPGERGGPGADGVAGPKGAGERSGPGA 517
QY 49 GPKGSGPGEAGRPGEAGLPGAKGLTGPSPGPDGKTGPPGAGQDGRFPFGARGQA 108
DB 518 GPKGSGPGEAGRPGEAGLPGAKGLTGPSPGPDGKTGPPGAGQDGRFPFGARGQA 577
QY 109 GYVGPFGPKGAAGEPKAGRGVPPGAVGAGKDGAGAGGPPGAGPAGERGEQGA 168
DB 578 GYVGPFGPKGAAGEPKAGRGVPPGAVGAGKDGAGAGGPPGAGPAGERGEQGA 637
QY 169 GSPGFGGLPGPAGPGEAGKPGGQVPGDLGAPGSPGAG 208
DB 638 GSPGFGGLPGPAGPGEAGKPGGQVPGDLGAPGSPGAG 677
RESULT 3
CALL RAT STANDARD; PRT; 671 AA.
ID _CALL1_RAT
AC P02454; P02455;
DT 21-JUL-1986 (Rel. 01, Created)
DT 01-FEB-1994 (Rel. 28, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Collagen alpha 1(I) chain (fragments).
GN COL1A1.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxID=101116;
RN [1]
RP SEQUENCE OF 1-19.
RX MEDLINE=69155173; PubMed=5777344;
RA Bornstein P.;
RT "Comparative sequence studies of rat skin and tendon collagen. II.
RT The absence of a short sequence at the amino terminus of the skin
RT alpha-1 chain.";
RL Biochemistry 8:63-71(1969).
[2]
RP SEQUENCE OF 5-19.
RX MEDLINE=67162288; PubMed=5337885;
RA Kang A.H., Bornstein P., Piez K.A.;
RT "The amino acid sequence of peptides from the cross-linking region of
RT rat skin collagen.";
RL Biochemistry 6:788-795(1967).
[3]
RP SEQUENCE OF 20-55.
RX MEDLINE=67165368; PubMed=4290711;
RA Bornstein P.;
RT "The incomplete hydroxylation of individual prolyl residues in
RT collagen.";

[2]
SEQUENCE OF 518-1128 FROM N.A.
MEDLINE=86137403; PubMed=3841523;
French B.T., Lee W.-H., Maul G.G.;
"Nucleotide sequence of a cDNA clone for mouse pro alpha 1(I)
collagen protein.";
Gene 39:311-312(1985).
[3]
SEQUENCE OF 735-1130 FROM N.A.
MEDLINE=8311374; PubMed=6298597;
Monson J.M., Friedman J., McCarthy B.J.;
"DNA sequence analysis of a mouse pro alpha 1(I) procollagen gene:
evidence for a mouse B1 element within the gene.";
Mol. Cell. Biol. 2:1362-1371(1982).
[4]
SEQUENCE OF 735-878 AND 1005-1058 FROM N.A.
MEDLINE=83157109; PubMed=6219867;
Monson J.M., McCarthy B.J.;
"Identification of a Balb/c mouse pro alpha 1(I) procollagen gene:
evidence for insertions or deletions in gene coding sequences.";
DNA 1:59-69(1981).
[5]
SEQUENCE OF 1442-1453 FROM N.A.
MEDLINE=88124276; PubMed=3340560;
Mooslehner K., Harbers K.;
"Two mRNAs of mouse pro alpha 1(I) collagen gene differ in the size
of the 3'-untranslated region.";
Nucleic Acids Res. 16:773-773(1988).
-!- FUNCTION: Type I collagen is a member of group I collagen
(fibrillar forming collagen).
-!- SUBUNIT: Trimers of one alpha 2(I) and two alpha 1(I) chains.
-!- TISSUE SPECIFICITY: Forms the fibrils of tendon, ligaments and
bones. In bones the fibrils are mineralized with calcium
hydroxyapatite.
-!- PTM: Prolines at the third position of the tripeptide repeating
unit (G-X-Y) are hydroxylated in some or all of the chains.
-!- SIMILARITY: Contains 1 WFEC domain.

This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
or send an email to license@isb-sib.ch).

EMBL; U08020; AAA88912.1; -
EMBL; X15896; CAA33904.1; -
EMBL; M14423; AAA37333.1; -
EMBL; M17491; AAA37334.1; -
EMBL; X06753; CAA29927.1; -
EMBL; K03036; AAA37332.1; -
EMBL; K03029; AAA37332.1; JOINED.
EMBL; K03030; AAA37332.1; JOINED.
EMBL; K03031; AAA37332.1; JOINED.
EMBL; K03032; AAA37332.1; JOINED.
EMBL; K03033; AAA37332.1; JOINED.
EMBL; K03034; AAA37332.1; JOINED.
EMBL; K03035; AAA37332.1; JOINED.
PIR; S57243; S21626
MGD; MGI:88467; Coll1a1.
DR InterPro; IPR008161; C1g_helix.
DR InterPro; IPR008160; Collagen.
DR InterPro; IPR008885; Fib_collagen_C.
DR InterPro; IPR002181; Fibrinogen_C.
DR InterPro; IPR001007; VWF_C.
DR Pfam; PF01410; COLFI_1.
DR Pfam; PF01391; Collagen_18.
DR ProDom; PD000007; C1g_helix; 1.
DR ProDom; PD002078; Fib_collagen_C; 1.
DR SMART; SM00038; COLFI; 1.
DR SMART; SM00214; VWC; 1.
DR PROSITE; PS01208; WFEC_1; 1.

DR PROSITE; PS50184; WFEC_2; 1.
KW Extracellular matrix; Connective tissue; Repeat; Hydroxylation;
Glycoprotein; Collagen; Signal.
FT SIGNAL 1 22
FT PROPEP 23 151 AMINO-TERMINAL PROPEPTIDE.
FT CHAIN 152 1207 COLLAGEN ALPHA 1(I) CHAIN.
FT PROPEP 1208 1453 CARBOXYL-TERMINAL PROPEPTIDE.
FT DOMAIN 29 87 WFEC.
FT DOMAIN 152 167 NONHELICAL REGION (N-TERMINAL).
FT DOMAIN 168 1181 TRIPLE-HELICAL REGION.
FT DOMAIN 1182 1207 NONHELICAL REGION (C-TERMINAL).
FT CARBOHYD 56 56 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 1354 1354 N-LINKED (GLCNAC. .) (POTENTIAL).
FT SITE 734 736 CELL ATTACHMENT SITE (POTENTIAL).
FT SITE 1082 1084 CELL ATTACHMENT SITE (POTENTIAL).
FT CONFLICT 1450 1450 A -> V (IN REF. 5).
SQ SEQUENCE 1453 AA; 137944 MW; 3B802E535DF81808 CRC64;

Query Match 94.3%; Score 1116; DB 1; Length 1453;
Best Local Similarity 90.5%; Pred. No. 1.5e-46;
Matches 199; Conservative 4; Mismatches 5; Indels 12; Gaps 1;

QY 1 GPP-----GEPCPTGLPGPGRGSGRGGPGADGVAGPKGAGRGSPGPA 48
|||
DB 447 GPPGAGEGKRGARGEPPGSLPGPGRGSGRGGPGADGVAGPKGSGERGAPGA 506
|||
QY 49 GPKGSGEAGRPCEAGLPCAKGLTGSPPSPGPDGKTGTPPGAGQDGRPGPPPGARGQA 108
|||
DB 507 GPKGSGEAGRPCEAGLPCAKGLTGSPPSPGPDGKTGTPPGAGQDGRPGPPPGARGQA 566
|||
QY 109 GVMGFPKGAGGPKAGRGVPPGPAVGAGKDXDGEAGAGPPGPPGAGRGFGQGA 168
|||
DB 567 GVMGFPKGAGGPKAGRGVPPGPAVGAGKDXDGEAGAGPPGPPGAGRGFGQGA 626
|||
QY 169 GSPGFGLPGPAGPCEAGKPGCEGQGVPGDLGAPGSGPAG 208
|||
DB 627 GSPGFGLPGPAGPCEAGKPGCEGQGVPGDLGAPGSGARG 666
|||

RESULT 5
CALL_CHICK STANDARD; PRT; 1453 AA.
AC P02457;
DT 21-JUL-1986 (Rel. 01, Created)
DT 01-OCT-1989 (Rel. 12, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Collagen alpha 1(I) chain precursor.
GN COL1A1.
OS Gallus gallus (Chicken).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Archosauria; Aves; Neognathae; Galliformes; Phasianidae;
OC Gallus.
CX NCBI_TaxID=9031;
RN [1]
RP SEQUENCE OF 1-153 FROM N.A.
RX MEDLINE=88056316; PubMed=3678834;
RA Finer M.H., Boedtker H., Doty P.;
RT "Construction and characterization of cDNA clones encoding the 5' end
of the chicken pro alpha 1(I) collagen mRNA.";
Gene 56:71-78(1987).
RN [2]
RP SEQUENCE OF 1-144 FROM N.A.
RX MEDLINE=88007542; PubMed=2820966;
RA Finer M.H., Aho S., Gerstenfeld L.C., Boedtker H., Doty P.;
RT "Unusual DNA sequences located within the promoter region and the
first intron of the chicken pro-alpha 1(I) collagen gene.";
J. Biol. Chem. 262:13323-13332(1987).
RN [3]
RP SEQUENCE OF 152-1187.
RX MEDLINE=82231995; PubMed=7093229;
RA Hignberger J.H., Corbett C., Dixit S.N., Yu W., Seyer J.M.,
Kang A.H., Gross J.;
RT "Amino acid sequence of chick skin collagen alpha 1(I)-CB8 and the

FT	DOMAIN	31	89	WFEC.
FT	MOD_RES	152	152	PYRROLIDONE CARBOXYLIC ACID.
FT	MOD_RES	254	254	HYDROXYLATION (POTENTIAL).
FT	MOD_RES	851	851	HYDROXYLATION (POTENTIAL).
FT	MOD_RES	1081	1081	HYDROXYLATION (POTENTIAL).
FT	MOD_RES	1097	1097	HYDROXYLATION (POTENTIAL).
FT	MOD_RES	1153	1153	HYDROXYLATION.
FT	CONFLICT	1187	1187	F -> L (IN REF. 5).
FT	CONFLICT	1441	1441	Q -> H (IN REF. 6).
FT	SEQUENCE	1453	AA; 137789	NW; 3BC6152134271F4D CRC64;
Query Match			90.1%;	Score 1067; DB 1; Length 1453;
Best Local Similarity			86.4%;	Pred. No. 3e-44;
Matches 190;			Conservative	4; Mismatches 14; Indels 12; Gaps 1;
QY	1	GPP-----	-----	-----GSPGPTGLPPGGRGGGSGRFFGADGVAGPKPAGERGSPGPA 48
DB	447	GPPGAGGEGKRGARGEPGAGL	PGPAGSRGAPGSRGFFGADGIAGPKPGERGSPGAV 506	
QY	49	GPKSGPEAGRPGAEGLPGAKGLT	SGSPGPDGKTPPGPGAGODRGPPGPGCARGQA 108	
DB	507	GPKSGPEAGRPGAEGLPGAKGLT	SGSGPDGKTPPGPGAGQDGRGPPGPGCARGQA 566	
QY	109	GVMGFPFGKGAAGPEPKAGERG	VGFGPGGAVGAGKDGAGAGQPPGPGPAGERGEOCPA 168	
DB	567	GVMGFPFGKGAAGPEPKAGERG	APGPPGAVGAGKDGAGAGQPPGPGPAGERGEOCPA 626	
QY	169	GSPFGQLPGPAGPPEAGKPG	EQGVPGDLGAPGSPGAG 208	
DB	627	GAPFGQLPGPAGPPEAGKPG	EQGVPGNAGAPGPAGARG 666	

RESULT 6

CA12_BOVIN	STANDARD;	PRT;	747 AA.
CA12_BOVIN	STANDARD;		
PO2459;	Q28070; Q9X124;		
21-JUL-1986	(Rel. 01, Created)		
30-MAY-2000	(Rel. 39, Last sequence update)		
15-MAR-2004	(Rel. 43, Last annotation update)		
Collagen alpha 1(II) chain precursor (Fragments).			
COL2A1			
Bos taurus (Bovine).			
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
Mammalia; Eutheria; Cetartiodactylia; Ruminantia; Pecora; Bovidae;			
Bovidae; Bovinae; Bos.			
NCBI_TaxID=9913;			
[1]			
SEQUENCE OF 1-15.			
TISSUE=Cartilage;			
MEDLINE=73258693; PubMed=4732855;			
Miller E.J., Lunde L.G.;			
"Isolation and Characterization of the cyanogen bromide peptides from			
the alpha 1(II) chain of bovine and human cartilage collagen.";			
Biochemistry 12:3153-3159(1973).			
[2]			
SEQUENCE OF 16-177.			
TISSUE=Cartilage;			
MEDLINE=76253504; PubMed=782511;			
Butler W.T., Miller E.J., Finch J.E. Jr.;			
"the covalent structure of cartilage collagen. Amino acid sequence of			
the NH2-terminal helical portion of the alpha 1 (II) chain.";			
Biochemistry 15:3000-3006(1976).			
[3]			
SEQUENCE OF 139-198.			
TISSUE=Cartilage;			
MEDLINE=77093864; PubMed=833147;			
Butler W.T., Finch J.E. Jr., Miller E.J.;			
"The covalent structure of cartilage collagen. Evidence for sequence			
heterogeneity of bovine alpha1(II) chains.";			
J. Biol. Chem. 252:639-643(1977).			
[4]			
SEQUENCE OF 139-417.			
TISSUE=Cartilage;			

RX MEDLINE=89231683; PubMed=2714276;
 RA Sever J.M., Hasty K.A., Kang A.H.;
 RT "Covalent structure of collagen. Amino acid sequence of an
 archaic cyanogen bromide peptide from type II collagen of
 bovine cartilage.";
 RL Eur. J. Biochem. 181:159-173 (1989).
 RN [5]
 RP SEQUENCE OF 418-492.
 RX MEDLINE=74163168; PubMed=4857180;
 RA Butler W.T., Miller E.J., Finch J.E. Jr., Inagami T.;
 RT "Homologous regions of collagen alpha(I) and alpha(II) chains:
 apparent clustering of variable and invariant amino acid residues.";
 RL Biochem. Biophys. Res. Commun. 57:190-195 (1974).
 RN [6]
 RP SEQUENCE OF 180-272 FROM N.A.
 RX TISSUE=Cartilage;
 RA Brand D.D., Myers L.K., Terato K., Whittington K.B., Stuart J.M.,
 Rosloniec E.F.;
 RT "Characterization of the T cell determinants in the induction of
 autoimmune arthritis by bovine alpha 1(II)-CB11 in H-2q mice.";
 RL J. Immunol. 152:3088-3097 (1994).
 RN [7]
 RP SEQUENCE OF 417-566 FROM N.A.
 RX TISSUE=Cartilage;
 RA Tang B., Chiang T.M., Brand D.D., Gumanovskaya M.L., Stuart J.M.,
 Kang A.H., Myers L.K.;
 RT "Molecular definition and characterization of recombinant bovine CB8
 and CB10: immunogenicity and arthritogenicity.";
 RL Clin. Immunol. 92:256-264 (1999).
 RN [8]
 RP SEQUENCE OF 567-747 FROM N.A.
 RX MEDLINE=85215651; PubMed=2582365;
 RA Sangiorgi F.O., Benson-Chanda V., de Wet W.J., Sobel M.E.,
 Ramirez F.;
 RT "Analysis of cDNA and genomic clones coding for the pro alpha 1 chain
 of calf type II collagen.";
 RL Nucleic Acids Res. 13:2815-2826 (1985).
 CC -!- FUNCTION: Collagen type II is specific for cartilaginous tissues.
 CC -!- SUBUNIT: Homotrimers of alpha 1(II) chains.
 CC -!- PTM: Proline residues at the third position in some or all of the
 repeating unit (G-X-Y) are hydroxylated in some or all of the
 chains. Hydroxylation on Pro-9 is involved in cross-linking.
 CC -!- PTM: O-linked glycans consist of Glc-Gal disaccharides bound to
 the oxygen atom of post-translationally added hydroxyl groups.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation -
 the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL; L28918; AAA30436.1; ..
 DR EMBL; AF138957; A2D42347.1; ..
 DR EMBL; X02420; CAB26269.1; ..
 DR PIR; A90363; CB06C.
 DR PIR; I45876; I45876.
 DR InterPro; IPR008161; Clg helix.
 DR InterPro; IPR008160; Collagen.
 DR InterPro; IPR000895; Fib collagen_C.
 DR InterPro; IPR001007; VFV C.
 DR ProDom; PD000007; Clg helix; 2.
 DR ProDom; PD002078; Fib collagen_C; 1.
 DR SMART; SM00038; COLFI; 1.
 DR PROSITE; PS01208; VFV C; PARTIAL.
 KW Extracellular matrix; Connective tissue; Repeat; Hydroxylation;
 Glycoprotein; Collagen.
 FT CHAIN 1 >566 COLLAGEN ALPHA 1(II) CHAIN.
 FT NON CONS 566 567
 FT PROPEP <567 747 CARBOXYL-TERMINAL PROPEPTIDE.

FT MOD RES 9 9
 FT MOD RES 102 102
 FT CARBOHYD 102 102
 FT MOD RES 114 114
 FT CARBOHYD 114 114
 FT MOD RES 123 123
 FT CARBOHYD 123 123
 FT MOD RES 189 189
 FT CARBOHYD 189 189
 FT MOD RES 423 423
 FT CARBOHYD 423 423
 FT MOD RES 435 435
 FT CARBOHYD 435 435
 FT VARIANT 143 143
 FT VARIANT 164 164
 FT CONFLICT 179 179
 FT CONFLICT 185 185
 FT CONFLICT 195 195
 FT CONFLICT 215 215
 FT CONFLICT 227 227
 FT CONFLICT 251 251
 FT CONFLICT 258 258
 FT CONFLICT 261 261
 FT CONFLICT 492 492
 SQ SEQUENCE 747 AA; 71329 MW; D0FCLD7CDICAF77C CRC64;
 Query Match 77.1%; Score 913; DB 1; Length 747;
 Best Local Similarity 78.4%; Pred. No. 3.6e-37;
 Matches 163; Conservative 6; Mismatches 39; Indels 0; Gaps 0;
 QY 1 GPPGEPGPTGLPQPGGPGSRGPGGAGVAGPKGPGAGRGSPGAGPKGSPGAGRP 60
 DB 307 GARGEPGGAGPAGPAGGPGGAGGPGGAGGPGGAGGPGGAGGPGGAGGPG 366
 QY 61 GEAGLPKAXGLTSGPSGPDGKTPGPGAGQDGRPPGPPGPPGPPGPPGPPGPPG 120
 DB 367 GEAGLPKAXGLTSGPSGPDGKTPGPGAGQDGRPPGPPGPPGPPGPPGPPG 426
 QY 121 GEPGKAGRGVPGPGGAVGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 180
 DB 427 GEPGKAGRGVPGPGGAVGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 486
 QY 181 GPFGEAGKPGGQGVPGDLCAGPGSPGAG 208
 DB 487 GPFGEAGKPGGQGVPGGAGAGPLVGRG 514
 RESULT 7
 CAL12 HUMAN STANDARD; PRT; 1418 AA.
 ID CAL12 HUMAN
 AC P02458;
 DT 21-JUL-1986 (Rel. 01, Created)
 DT 01-JAN-1990 (Rel. 13, Last sequence update)
 DT 15-MAR-2004 (Rel. 43, Last annotation update)
 DE Collagen alpha 1(II) chain precursor [Contains: Chondrocalcin].
 GN COL2A1.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=90067946; PubMed=2587267;
 RA Su M.W., Lee B., Ramirez F., Machado M., Horton W.;
 RT "Nucleotide sequence of the full length cDNA encoding for human type
 II procollagen.";
 RL Nucleic Acids Res. 17:9473-9473 (1989).
 RN [2]
 RP SEQUENCE OF 1-28 FROM N.A.
 RX MEDLINE=87031574; PubMed=3021582;
 RA Nunez A.M., Kohno K., Martin G.R., Yamada Y.;
 RT "Promoter region of the human pro-alpha 1(II)-collagen gene.";
 RL Gene 44:11-16 (1986).

RN [3] SEQUENCE OF 432-1145 FROM N.A.
 RA Ramirez F.;
 RL Submitted (DEC-1988) to the EMBL/GenBank/DBJ databases.
 RN [4]
 RP SEQUENCE OF 963-1418 FROM N.A.
 RX MEDLINE=85190334; PubMed=3857598;
 RA Cheah K.S.E., Stoker N.G., Griffin J.R., Grosveld F.G., Solomon E.;
 RT "Identification and characterization of the human type II collagen
 gene (COL2A1).";
 RL Proc. Natl. Acad. Sci. U.S.A. 82:2555-2559(1985).
 RN [5]
 RP SEQUENCE OF 1120-1398 FROM N.A.
 RX MEDLINE=85306861; PubMed=3840017;
 RA Elina K., Maekelae J.K., Vuorio T., Kauppinen S., Knowles J.,
 RA Vuorio E.;
 RT "Construction and identification of a cDNA clone for human type II
 procollagen mRNA.";
 RL Biochem. J. 229:183-188(1985).
 RN [6]
 RP SEQUENCE OF 1106-1418 FROM N.A.
 RX MEDLINE=88067771; PubMed=2825137;
 RA Elina K., Vuorio T., Vuorio E.;
 RT "Determination of the single polyadenylation site of the human pro
 alpha 1(II) collagen gene.";
 RL Nucleic Acids Res. 15:9499-9504(1987).
 RN [7]
 RP SEQUENCE OF 1227-1289 FROM N.A.
 RX MEDLINE=86104139; PubMed=3002437;
 RA Nunez A.M., Francomano C., Young M.F., Martin G.R., Yamada Y.;
 RT "Isolation and partial characterization of genomic clones coding for
 a human pro-alpha 1 (II) collagen chain and demonstration of
 restriction fragment length polymorphism at the 3' end of the gene.";
 RL Biochemistry 24:6343-6348(1985).
 RN [8]
 RP SEQUENCE OF 1176-1226 FROM N.A.
 RX MEDLINE=84118798; PubMed=6320112;
 RA Strom C.M., Upholt W.B.;
 RT "Isolation and characterization of genomic clones corresponding to
 the human type II procollagen gene.";
 RL Nucleic Acids Res. 12:1025-1038(1984).
 RN [9]
 RP SEQUENCE OF 35-167 FROM N.A.
 RX MEDLINE=89233138; PubMed=2714801;
 RA Su M.W., Benson-Chanda V., Vissing H., Ramirez F.;
 RT "Organization of the exons coding for pro alpha 1(II) collagen N-
 propeptide confirms a distinct evolutionary history of this domain of
 the fibrillar collagen genes.";
 RL Genomics 4:438-441(1989).
 RN [10]
 RP REVIEW ON VARIANTS.
 RX MEDLINE=91184577; PubMed=2010058;
 RA Kuivaniemi H., Tromp G., Prockop D.J.;
 RT "Mutations in collagen genes: causes of rare and some common diseases
 in humans.";
 RL PASEB J. 5:2052-2060(1991).
 RN [11]
 RP REVIEW ON VARIANTS.
 RX MEDLINE=97255959; PubMed=9101290;
 RA Kuivaniemi H., Tromp G., Prockop D.J.;
 RT "Mutations in fibrillar collagens (types I, II, III, and XI), fibril-
 associated collagen (type IX), and network-forming collagen (type X)
 cause a spectrum of diseases of bone, cartilage, and blood vessels.";
 RL Hum. Mutat. 9:300-315(1997).
 RN [12]
 RP VARIANT SER-1074.
 RX MEDLINE=30036909; PubMed=2572591;
 RA Vissing H., D'Alessio M., Lee B., Ramirez F., Godfrey M.,
 RA Hollister D.W.;
 RT "Glycine to serine substitution in the triple helical domain of pro-
 alpha 1(III) collagen results in a lethal perinatal form of short-
 limbed dwarfism.";
 RL J. Biol. Chem. 264:18265-18267(1989).
 RN [13]
 RP VARIANT SEDC 1095-GLY--TYR-1330 DEL.
 RX MEDLINE=89266907; PubMed=2543071;
 RA Lee B., Vissing H., Ramirez F., Rogers D., Rimo D.J.;
 RT "Identification of the molecular defect in a family with
 spondyloepiphyseal dysplasia.";
 RL Science 244:978-980(1989).
 RN [14]
 RP VARIANT OSTEOARTHRITIS CYS-650.
 RX MEDLINE=90370826; PubMed=1975693;
 RA Ala-Kokko L., Baldwin C.T., Moskowitz R.W., Prockop D.J.;
 RT "Single base mutation in the type II procollagen gene (COL2A1) as a
 cause of primary osteoarthritis associated with a mild
 chondrodysplasia.";
 RL Proc. Natl. Acad. Sci. U.S.A. 87:6565-6568(1990).
 RN [15]
 RP VARIANT OI-IV VAL-717.
 RX MEDLINE=91291136; PubMed=2064612;
 RA Bateman J.F., Hannagan M., Chan D., Cole W.G.;
 RT "Characterization of a type I collagen alpha 2(I) glycine-586 to
 valine substitution in osteogenesis imperfecta type IV. Detection of
 the mutation and prenatal diagnosis by a chemical cleavage method.";
 RL Biochem. J. 276:765-770(1991).
 RN [16]
 RP VARIANT OSTEOARTHRITIS CYS-650.
 RX MEDLINE=91086471; PubMed=1985108;
 RA Eyre D.R., Weis M.A., Moskowitz R.W.;
 RT "Cartilage expression of a type II collagen mutation in an inherited
 form of osteoarthritis associated with a mild chondrodysplasia.";
 RL J. Clin. Invest. 87:357-361(1991).
 RN [17]
 RP VARIANT HYPOCHONDROGENESIS GLU-984.
 RX MEDLINE=93054548; PubMed=1429602;
 RA Bogart R., Tiller G.E., Wies M.A., Gruber H.E., Rimo D.L.,
 RA Cohn D.H., Eyre D.R.;
 RT "An amino acid substitution (Gly983-->Glu) in the collagen alpha
 1(II) chain produces hypochondrogenesis.";
 RL J. Biol. Chem. 267:22523-22526(1992).
 RN [18]
 RP VARIANT HYPOCHONDROGENESIS SER-705.
 RX MEDLINE=92262484; PubMed=1374506;
 RA Horton W.A., Machado M.A., Ellard J., Campbell D., Bartley J.,
 RA Ramirez F., Vitale E., Lee B.;
 RT "Characterization of a type II collagen gene (COL2A1) mutation
 identified in cultured chondrocytes from human hypochondrogenesis.";
 RL Proc. Natl. Acad. Sci. U.S.A. 89:4583-4587(1992).
 RN [19]
 RP VARIANT WS-II ASP-198.
 RX MEDLINE=93304428; PubMed=8317498;
 RA Koerkoe J., Ritvaniemi P., Haataja L., Kaeaeiaieinen H.,
 RA Kivirikko K.I., Prockop D.J., Ala-Kokko L.;
 RT "Mutation in type II procollagen (COL2A1) that substitutes aspartate
 for glycine alpha 1-67 and that causes cataracts and retinal
 detachment: evidence for molecular heterogeneity in the Wagner
 syndrome and the Stickler syndrome (arthro-ophthalmopathy).";
 RL Am. J. Hum. Genet. 53:55-61(1993).
 RN [20]
 RP VARIANT SEMD CYS-840.
 RA Tiller G.E., Weis M.A., Lachman R.S., Cohn D.H., Rimo D.L.,
 RA Eyre D.R.;
 RT "A dominant mutation in the type II collagen gene (COL2A1) produces
 spondyloepimephal dysplasia (SEMD), Strudwick type.";
 RL Am. J. Hum. Genet. 53:A209-A209(1993).
 RN [21]
 RP VARIANT OSTEOARTHRITIS CYS-650.
 RX MEDLINE=93282819; PubMed=8507190;
 RA Holderbaum D., Malemud C.J., Moskowitz R.W., Haggi T.M.;
 RT "Human cartilage from late stage familial osteoarthritis transcribes
 type II collagen mRNA encoding a cysteine in position 519.";
 RL Biochem. Biophys. Res. Commun. 192:1169-1174(1993).
 RN [22]
 RP VARIANT SEMD ARG-285.
 RX MEDLINE=93252400; PubMed=8486375;


```

QY 181 GPPGAGKPGEGVPGDLGAPGSPGAG 208
DB 644 GPPGEGGKGQDGPGEAGAPGLVGRG 671

RESULT 9
CALL_BOVIN
ID CALL_BOVIN STANDARD; PRT; 779 AA.
AC P02453;
DT 21-JUL-1986 (Rel. 01, Created)
DT 01-FEB-1994 (Rel. 28, Last sequence update)
DT 15-MAR-2004 (Rel. 43, Last annotation update)
DE Collagen alpha 1(I) chain (Fragments).
GN COL1A1
OS Bos taurus (Bovine).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
OC Bovidae; Bovinae; Bos.
OC NCBI_TaxID=9913;
RN [1]
RP SEQUENCE OF 1-19.
RX MEDLINE=72255334; PubMed=4115172;
RA Rautenberg J., Timpl R., Furthmayr H.;
RT "Structural characterization of N-terminal antigenic determinants in
RT calf and human collagen."
RL Eur. J. Biochem. 27:231-237(1972).
RN [2]
RP SEQUENCE OF 20-145.
RX MEDLINE=76022320; PubMed=1164916;
RA Fietzek P.P., Kuehn K.;
RT "The covalent structure of collagen: amino-acid sequence of the
RT cyanogen-bromide peptides alpha-1-CB2, alpha-1-CB4 and alpha-1-CB5
RT from calf-skin collagen."
RL Eur. J. Biochem. 52:77-82(1975).
RN [3]
RP SEQUENCE OF 146-294.
RX MEDLINE=73049499; PubMed=4673951;
RA Fietzek P.P., Wendt P., Kell I., Kuehn K.;
RT "The covalent structure of collagen: amino acid sequence of alpha-1-
RT CB3 from calf skin collagen."
RL FEBS Lett. 26:74-76(1972).
RN [4]
RP SEQUENCE OF 295-562.
RX MEDLINE=74086118; PubMed=4359390;
RA Fietzek P.P., Rexrodt F.W., Hopper K.E., Kuehn K.;
RT "The covalent structure of collagen. 2. The amino-acid sequence of
RT alpha-1-CB7 from calf-skin collagen."
RL Eur. J. Biochem. 38:396-400(1973).
RN [5]
RP SEQUENCE OF 563-675.
RX MEDLINE=73042276; PubMed=4343808;
RA Wendt P., Mark K.V.D., Rexrodt F., Kuehn K.;
RT "The covalent structure of collagen. The amino-acid sequence of the
RT 112-residues. Amino-terminal part of peptide alpha-1-CB6 from calf-
RT skin collagen."
RL Eur. J. Biochem. 30:169-183(1972).
RN [6]
RP SEQUENCE OF 676-751.
RX MEDLINE=73042275; PubMed=4343807;
RA Fietzek P.P., Rexrodt F.W., Wendt P., Stark M., Kuehn K.;
RT "The covalent structure of collagen. Amino-acid sequence of peptide
RT alpha-1-CB6-C2."
RL Eur. J. Biochem. 30:163-168(1972).
CC -1- FUNCTION: Type I collagen is a member of group I collagen
CC (fibrillar forming collagen).
CC -1- SUBUNIT: Trimers of one alpha 2(I) and two alpha 1(I) chains.
CC -1- TISSUE SPECIFICITY: Forms the fibrils of tendon, ligaments and
CC bones. In bones the fibrils are mineralized with calcium
CC hydroxyapatite.
CC -1- PMW: Proline residues at the third position of the tripeptide
CC repeating unit (G-X-Y) are hydroxylated in some or all of the
CC chains. Pro-726 is the only 3-hydroxypro and the only hydroxylated

```

SEQUENCE OF 29-96; 332-397; 431-484; 503-535 AND 869-976 FROM N.A.
 MEDLINE=84270696; PubMed=6547770;
 Yanada Y., Liang G., Mudryj M., Obici S., de Crombrughe B.;
 "Conservation of the sizes for one but not another class of exons in
 two chick collagen genes.";
 Nature 310:333-337(1984).
 [3]
 SEQUENCE OF 977-1262 FROM N.A.
 MEDLINE=83220816; PubMed=6856474;
 Yanada Y., Kuhn K., de Crombrughe B.;
 "A conserved nucleotide sequence, coding for a segment of the C-
 propeptide, is found at the same location in different collagen
 genes.";
 Nucleic Acids Res. 11:2733-2744(1983).
 CC -!- FUNCTION: Collagen type III occurs in most soft connective tissues
 along with type I collagen.
 CC -!- SUBUNIT: Trimers of identical alpha 1(III) chains. The chains are
 linked to each other by interchain disulfide bonds. Trimers are
 also cross-linked via hydroxylsines.
 CC -!- PTM: Prolines at the third position of the tripeptide repeating
 unit (G-X-Y) are hydroxylated in some or all of the chains.
 CC -!- SIMILARITY: Contains 1 VWFC domain.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation -
 the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 or send an email to license@isb-sib.ch).

 DR EMBL; U07973; AA83407.1; -
 DR EMBL; X00822; CAB52866.1; -
 DR EMBL; X00823; CAB52866.1; JOINED.
 DR EMBL; X00826; CAA25397.1; ALT_SEQ.
 DR EMBL; X00825; CAA25397.1; JOINED.
 DR EMBL; X00827; CAA25398.1; -
 DR EMBL; X00828; CAA25399.1; -
 DR EMBL; X00830; CAA25401.1; -
 DR EMBL; X00831; CAA25402.1; -
 DR EMBL; X02302; AAD15299.1; -
 DR EMBL; K02301; AAD15298.1; -
 DR EMBL; M36662; AAA18519.1; ALT_SEQ.
 DR PIR; A05269; A05269.
 DR PIR; I50694; I50694.
 DR InterPro; IPR008161; C1g helix.
 DR InterPro; IPR008160; Collagen.
 DR InterPro; IPR000885; Fib collagen C.
 DR InterPro; IPR002181; Fibrinogen C.
 DR InterPro; IPR001007; VWF C.
 DR ProDom; PD000007; C1g helix; 1.
 DR ProDom; PD002078; Fib collagen_C; 1.
 DR SMART; SM00038; COLF1; 1.
 DR SMART; SM00214; VWC; 1.
 DR PROSITE; PS01208; VWC; 1.
 DR PROSITE; PS0184; VWC2; 1.
 KW Extracellular matrix; Connective tissue; Repeat; Hydroxylation;
 Glycoprotein; Collagen; Signal.
 FT SIGNAL 1 23
 FT PROPEP 24 144
 FT AMINO-TERMINAL PROPEPTIDE (BY
 FT SIMILARITY).
 FT CHAIN 145 1003
 FT COLLAGEN ALPHA 1(III) CHAIN.
 FT PROPEP 1004 1262
 FT CARBOXYL-TERMINAL PROPEPTIDE (BY
 FT SIMILARITY).
 FT VWF.
 FT DOMAIN 29 88
 FT DOMAIN 145 164
 FT NONHELICAL REGION (N-TERMINAL) (BY
 FT SIMILARITY).
 FT DOMAIN 165 994
 FT TRIPLE-HELICAL REGION (BY SIMILARITY).
 FT DOMAIN 995 1003
 FT NONHELICAL REGION (C-TERMINAL) (BY
 FT SIMILARITY).
 FT NON_CONS 886 887
 FT NON_CONS 922 923
 FT DISULFID 994 994
 FT INTERCHAIN (BY SIMILARITY).

FT DISULFID 995 995 INTERCHAIN (BY SIMILARITY).
 FT MOD_RES 262 262 HYDROXYLATION (BY SIMILARITY).
 FT MOD_RES 283 283 HYDROXYLATION (BY SIMILARITY).
 FT MOD_RES 859 859 HYDROXYLATION (BY SIMILARITY).
 FT CARBOHYD 1163 1163 N-LINKED (GLCNAC...) (POTENTIAL).
 FT CONFLICT 96 96 E -> K (IN REF. 2).
 FT CONFLICT 1132 1132 F -> S (IN REF. 3).
 SQ SEQUENCE 1262 AA; 121249 MW; 96ABE7B2E9DEB43D CRC64;
 Query Match 67.1%; Score 794.5; DB 1; Length 1262;
 Best Local Similarity 65.9%; Pred. No. 2e-31;
 Matches 143; Conservative 13; Mismatches 52; Indels 9; Gaps 1;
 Qy 1 GPP-----GEPDPTGLPDPGPGGPGSRGPGFADGAGPGPAGERSGPGAGPK 51
 Db 458 GPGEGEKGANGEPGQNGVPTGGERGSGPGRGLPGNGLPGEKGPAGERSGPGPGPS 517
 Qy 52 GSPEAGRPEAGLPGAKGLTSPGSPGDKTGPAGQDGRPPGPPGARGQAGYM 111
 Db 518 GPAGDGGDGGCLFGMRGLPGIPGSPGSDGKPGPGNGQGEGRSGPPGAGPRGQPGYM 577
 Qy 112 GPPGPKGAAGEPKAGRGVPPGPAAGVAGKDGAGAGGPPGPPGAGPAGEREQPGASP 171
 Db 578 GFPGPKNEGAPGKNGERGPGGPPGTPGPAKNGDVGLFPGPPGAGPADGRGEPGSGSP 637
 Qy 172 GFQGLPAPGPGCEAGKPGEGQGVGDLGAPGSGPAG 208
 Db 638 GLOGLPGPGPAGENGKPGEPGKGDIGGPGPGPKG 674
 RESULT 11
 CA25 HUMAN
 ID CA25 HUMAN STANDARD; PRT; 1496 AA.
 AC P05997;
 DT 01-APR-1988 (Rel. 07, Created)
 DT 01-JAN-1990 (Rel. 13, Last sequence update)
 DT 10-OCT-2003 (Rel. 42, Last annotation update)
 DE Collagen alpha 2(V) chain precursor.
 GN COL5A2
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE OF 1-463 FROM N.A.
 RX MEDLINE=89123368; PubMed=2914927;
 RA Woodbury D., Benson-Chanda V., Ramirez F.;
 "Amino-terminal propeptide of human pro-alpha 2(V) collagen conforms
 to the structural criteria of a fibrillar procollagen molecule.";
 J. Biol. Chem. 264:2735-2738(1989).
 RN [2]
 RP SEQUENCE OF 398-1496 FROM N.A.
 RX MEDLINE=87146331; PubMed=3029669;
 RA Weil D., Bernard M.P., Gargano S., Ramirez F.;
 "The pro alpha 2(V) collagen gene is evolutionarily related to the
 major fibrillar-forming collagens";
 Nucleic Acids Res. 15:181-198(1987).
 RN [3]
 RP SEQUENCE OF 1227-1496 FROM N.A.
 RX MEDLINE=85289337; PubMed=2411731;
 RA Myers J.C., Loidl H.R., Seyer J.M., Dion A.S.;
 "Complete primary structure of the human alpha 2 type V procollagen
 COOH-terminal propeptide";
 J. Biol. Chem. 260:11216-11222(1985).
 RN [4]
 RP SEQUENCE OF 1449-1496 FROM N.A.
 RX MEDLINE=89138450; PubMed=3224983;
 RA Tsiouras P., Schwartz R.C., Liddell A.C., Salkeld C.S., Weil D.,
 Ramirez F.;
 "Genetic distance of two fibrillar collagen loci, COL3A1 and COL5A2,
 located on the long arm of human chromosome 2.";
 Genomics 3:275-277(1988).
 RN [5]

SEQUENCE OF 208-227.
 RC TISSUE=Placenta;
 RA MEDLINE=92239022; PubMed=1571108;
 RX Mann K.;
 RT "Isolation of the alpha 3-chain of human type V collagen and
 RL characterization by partial sequencing";
 RN Biol. Chem. Hoppe-Seyler 373:63-75(1992).
 [6]
 RP SEQUENCE OF 288-297 AND 606-617.
 RC TISSUE=Bone;
 RA MEDLINE=94237164; PubMed=8181482;
 RX Moradi-Ameli M., Rousseau J.C., Klemm J.P., Champlaud M.F.,
 RA Bouillon M.M., Bernillon J., Wallach J.M., van der Rest M.;
 RT "Diversity in the processing events at the N-terminus of type-V
 RL collagen";
 RN Eur. J. Biochem. 221:987-995(1994).
 [7]
 RP SEQUENCE OF 1226-1227.
 RC TISSUE=Collagen;
 RA MEDLINE=98087576; PubMed=9425231;
 RX Michalikova K., Susic M., Willing M.C., Wenstrup R.J., Cole W.G.;
 RA "Mutations of the alpha2(V) chain of type V collagen impair matrix
 RT assembly and produce Ehlers-Danlos syndrome type I";
 RN Hum. Mol. Genet. 7:249-255(1998).
 [8]
 RP VARIANT EDS-II ARG-960.
 RC MEDLINE=98455031; PubMed=9783710;
 RA Richards A.J., Martin S., Nicholls A.C., Harrison J.B., Pope F.M.,
 RA Burrows N.P.;
 RT "A single base mutation in COL5A2 causes Ehlers-Danlos syndrome type
 RL II";
 RN J. Med. Genet. 35:846-848(1998).
 CC -!- FUNCTION: Type V collagen is a member of group I collagen
 CC (fibrillar forming collagen). It is a minor connective tissue
 CC component of nearly ubiquitous distribution. Type V collagen binds
 CC to DNA, heparan sulfate, thrombospondin, heparin, and insulin.
 CC -!- SUBUNIT: Trimers of two alpha 1(V) and one alpha 2(V) chains in
 CC most tissues and trimers of one alpha 1(V), one alpha 2(V), and
 CC one alpha 3(V) chains in placenta.
 CC -!- PTM: Prolines at the third position of the tripeptide repeating
 CC unit (G-X-Y) are hydroxylated in some or all of the chains.
 CC -!- DISEASE: Defects in COL5A2 are a cause of Ehlers-Danlos syndrome
 CC graveis. EDS-I is an autosomal dominant connective-tissue disorder
 CC characterized by loose-jointedness and fragile, velvety,
 CC stretchable, bruisable skin that heals with peculiar 'cigarette-
 CC paper' scars. Inheritance is autosomal dominant.
 CC -!- DISEASE: Defects in COL5A2 are a cause of Ehlers-Danlos
 CC syndrome mitis. Inheritance is autosomal dominant.
 CC -!- SIMILARITY: Contains 1 WFEC domain.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 CC EMBL: J04478; AAA51859.1; -;
 CC EMBL: X04758; CAA28454.1; -;
 CC EMBL: M11718; AAA52058.1; -;
 CC PIR: A31427; CGHU2V
 CC PDB: 1A9A; 18-NOV-98.
 CC Genew; HGNC:2210; COL5A2.
 CC MIM: 120190; -;
 CC MIM: 130000; -;
 CC MIM: 130010; -;
 CC GO: GO:0005588; C:collagen type V; TAS.
 CC GO: GO:0008151; P:cell growth and/or maintenance; TAS.
 CC InterPro: IPR008161; C1g_helix.
 CC InterPro: IPR008160; Collagen.
 CC InterPro: IPR000885; Fib_collagen_C.

DR InterPro: IPR002181; Fibrinogen_C.
 DR InterPro: IPR001007; VWF_C.
 DR Pfam: PF01410; COLFI; 1.
 DR Pfam: PF01391; Collagen; 18.
 DR Pfam: PF00093; vwc; 1.
 DR ProDom: PD000007; C1g_helix; 5.
 DR ProDom: PD002078; Fib_collagen_C; 1.
 DR SMART: SM00038; COLFI; 1.
 DR SMART: SM00214; VMC; 1.
 DR PROSITE: PS01208; VWF; 1; 1.
 DR PROSITE: PS0184; VWF; 2; 1.
 KW Extracellular matrix; Connective tissue; Repeat; Hydroxylation;
 KW Glycoprotein; Collagen; Signal; Ehlers-Danlos syndrome;
 KW Disease mutation; 3D-structure.
 FT SIGNAL 1 26
 FT CHAIN 27 1226 COLLAGEN ALPHA 2(V) CHAIN.
 FT PROPEP 1227 1496 CARBOXYL-TERMINAL PROPEPTIDE.
 FT DOMAIN 39 97 VWF.
 FT MOD_RES 290 290 HYDROXYLATION.
 FT MOD_RES 293 293 HYDROXYLATION.
 FT MOD_RES 296 296 HYDROXYLATION.
 FT MOD_RES 608 608 HYDROXYLATION.
 FT MOD_RES 614 614 HYDROXYLATION.
 FT VARIANT 960 960 G -> R (in EDS-II).
 FT CONFLICT 292 292 A -> P (IN REF. 6).
 FT CONFLICT 1418 1418 K -> T (IN REF. 3).
 FT CONFLICT 1438 1438 F -> S (IN REF. 3).
 FT CONFLICT 1460 1460 E -> Q (IN REF. 4).
 FT CONFLICT 1496 1496 V -> A (IN REF. 4).
 SQ SEQUENCE 1496 AA; 144720 MW; 82827C17A8644F5A CRC64;
 Query Match 66.4%; Score 786; DB 1; Length 1496;
 Best Local Similarity 67.3%; Pred. No. 5.6e-31;
 Matches 140; Conservative 15; Mismatches 53; Indels 0; Gaps 0;
 QY 1 GPPGPGPTGLPFGPGRGGPGRGPGADGAGVAGPKGAGRGSPGAPGKSGPGEAGRP 60
 Db 501 GPRGPGTGLPFGPGRGAPGRGPGDGLPGKGAQGERGPGVSGSPKSGQDPRP 560
 QY 61 GAGLPGAKLTGSPGSPGPKTTPPGAGQDRPPGPPGARGQAGVWGPFGPKGAA 120
 Db 561 GEPGLPARGLTGNFVGQGPGLGAPGEDRPPGPGSIGIKGPGTMTGLFPGKGSN 620
 QY 121 GPPGKAGRGVPPGPGVAGPGAGKGAQGPAGRGAGRGAGRGAGRGAGRGAGRG 180
 Db 621 GPPGKAGRGVPPGPGVAGPGAGKGAQGPAGRGAGRGAGRGAGRGAGRGAGRG 680
 QY 181 GPPGKAGRGVPPGPGVAGPGAGKGAQGPAGRGAGRGAGRGAGRGAGRGAGRG 208
 Db 681 GPPGKAGRGVPPGPGVAGPGAGKGAQGPAGRGAGRGAGRGAGRGAGRGAGRG 708
 RESULT 12
 ID CAL3_HUMAN STANDARD; PRT; 1466 AA.
 AC P02461; Q15112;
 DT 21-JUL-1986 (Rel. 01, Created)
 DT 01-JAN-1990 (Rel. 13, Last sequence update)
 DT 10-OCT-2003 (Rel. 42, Last annotation update)
 DE Collagen alpha 1 (III) chain precursor.
 GN COL3A1.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 RN [1]
 RX NCBI_TaxID=9606;
 RP SEQUENCE FROM N.A.
 RC TISSUE=Skin fibroblast;
 RX MEDLINE=89350838; PubMed=2764886;
 RA Ala-Kokko L., Kontusaari S., Baldwin C.T., Kuivaniemi H.,
 RA Prockop D.J.;
 RT "Structure of cDNA clones coding for the entire prepro alpha 1 (III)

RT chain of human type III procollagen. Differences in protein structure
 RT from type I procollagen and conservation of codon preferences.";
 RL Biochem. J. 260:509-516(1989).
 RP [2]
 RP SEQUENCE OF 149-1225 FROM N.A.
 RX MEDLINE=89386015; PubMed=2780304;
 RA Janeczko R.A., Ramirez F.;
 RT "Nucleotide and amino acid sequences of the entire human alpha 1
 RT (III) collagen.";
 RL Nucleic Acids Res. 17:6742-6742(1989).
 RP [3]
 RP SEQUENCE OF 168-398.
 RX MEDLINE=77134724; PubMed=557335;
 RA Seyer J.M., Kang A.H.;
 RT "Covalent structure of collagen: amino acid sequence of cyanogen
 RT bromide peptides from the amino-terminal segment of type III collagen
 RT of human liver.";
 RL Biochemistry 16:1158-1164(1977).
 RP [4]
 RP REVISIONS.
 RA Seyer J.M.;
 RL Submitted (DEC-1977) to the PIR data bank.
 RP [5]
 RP SEQUENCE OF 399-727.
 RX MEDLINE=7900343; PubMed=687591;
 RA Seyer J.M., Kang A.H.;
 RT "Covalent structure of collagen: amino acid sequence of five
 RT consecutive CNBr peptides from type III collagen of human liver.";
 RL Biochemistry 17:3404-3411(1978).
 RP [6]
 RP SEQUENCE OF 728-964.
 RX MEDLINE=80198282; PubMed=6246925;
 RA Seyer J.M., Mainardi C., Kang A.H.;
 RT "Covalent structure of collagen: amino acid sequence of alpha 1
 RT (III)-CB5 from type III collagen of human liver.";
 RL Biochemistry 19:1583-1589(1980).
 RP [7]
 RP SEQUENCE OF 950-1466 FROM N.A.
 RX MEDLINE=88189827; PubMed=3357782;
 RA Mankoo B.S., Dalgleish R.;
 RT "Human pro alpha 1(III) collagen: cDNA sequence for the 3' end.";
 RL Nucleic Acids Res. 16:2337-2337(1988).
 RP [8]
 RP REVISION TO 1184.
 RX MEDLINE=89098346; PubMed=3211760;
 RA Molyneux K., Dalgleish R.;
 RT "Human type III collagen 'variant' is a cDNA cloning artefact.";
 RL Nucleic Acids Res. 16:11833-11833(1988).
 RP [9]
 RP SEQUENCE OF 1065-1466 FROM N.A.
 RX MEDLINE=85087944; PubMed=6096827;
 RA Loidl H.R., Brinker J.M., May M., Pihlajaniemi T., Morrow S.,
 RT Rosenbloom J., Myers J.C.;
 RT "Molecular cloning and carboxyl-propeptide analysis of human type III
 RT procollagen.";
 RL Nucleic Acids Res. 12:9383-9394(1984).
 RP [10]
 RP SEQUENCE OF 965-1200.
 RX MEDLINE=81208139; PubMed=7016180;
 RA Seyer J.M., Kang A.H.;
 RT "Covalent structure of collagen: amino acid sequence of alpha
 RT 1(III)-CB9 from type III collagen of human liver.";
 RL Biochemistry 20:2621-2627(1981).
 RP [11]
 RP SEQUENCE OF 1176-1466 FROM N.A.
 RX MEDLINE=85157600; PubMed=2579949;
 RA Chu M.-L., Weil D., de Wet W.J., Bernard M.P., Sippola M., Ramirez F.;
 RT "Isolation of cDNA and genomic clones encoding human pro-alpha 1
 RT (III) collagen. Partial characterization of the 3' end region of the
 RT gene.";
 RL J. Biol. Chem. 260:4357-4363(1985).
 RP [12]
 RP SEQUENCE OF 1161-1200 FROM N.A.

RX MEDLINE=86187804; PubMed=3754462;
 RA Misulin M., Dalgleish R., Kluge-Beckerman B., Rennard S.I.,
 RA Tolstoshev P., Brantly M., Crystal R.G.;
 RT "Human type III collagen gene expression is coordinately modulated
 RT with the type I collagen genes during fibroblast growth.";
 RL Biochemistry 25:1408-1413(1986).
 RP [13]
 RP SEQUENCE OF 1-170 FROM N.A.
 RX TISSUE=Placenta;
 RA MEDLINE=88303360; PubMed=3405773;
 RA Toman D., Ricca G., de Crombrughe B.;
 RT "Nucleotide sequence of a cDNA coding for the amino-terminal region
 RT of human pro-alpha 1(III) collagen.";
 RL Nucleic Acids Res. 16:7201-7201(1988).
 RP [14]
 RP SEQUENCE OF 1-176 FROM N.A.
 RX MEDLINE=89378752; PubMed=2777083;
 RA Benson-Chanda V., Su M.W., Weil D., Chu M.-L., Ramirez F.;
 RT "Cloning and analysis of the 5' portion of the human type-III
 RT procollagen gene (COL3A1).";
 RL Gene 78:255-265(1989).
 RP [15]
 RP REVIEW ON VARIANTS.
 RX MEDLINE=97255959; PubMed=9101290;
 RA Kuivaniemi H., Tromp G., Prockop D.J.;
 RT "Mutations in fibrillar collagens (types I, II, III, and XI), fibril-
 RT associated collagen (type IX), and network-forming collagen (type X)
 RT cause a spectrum of diseases of bone, cartilage, and blood vessels.";
 RL Hum. Mutat. 9:300-315(1997).
 RP [16]
 RP VARIANT AORTIC ANEURYSM ARG-303, AND VARIANT THR-668.
 RX MEDLINE=93293988; PubMed=8514866;
 RA Tromp G., Wu Y., Prockop D.J., Madhathari S.L., Kleinert C.,
 RA Earley J.J., Zhuang J., Noergaard O., Darling R.C., Abbott W.M.,
 RA Cole C.W., Jaakkola P., Ryyanen M., Pearce W.H., Yao J.S.T., S.A.,
 RA Majamaa K., Smullen S.N., Gatalica Z., Ferrell R.E., Jimenez S.A.,
 RA Jackson C.E., Michels V.V., Kaye M., Kuivaniemi H.;
 RT "Sequencing of cDNA from 50 unrelated patients reveals that mutations
 RT in the triple-helical domain of type III procollagen are an
 RT infrequent cause of aortic aneurysms.";
 RL J. Clin. Invest. 91:2539-2545(1993).
 RP [17]
 RP VARIANT THR-698.
 RX MEDLINE=91045136; PubMed=2235526;
 RA Zafarullah K., Kleinert C., Tromp G., Kuivaniemi H., Kontusaari S.,
 RA Wu Y., Ganguly A., Prockop D.J.;
 RT "G to A polymorphism in exon 31 of the COL3A1 gene.";
 RL Nucleic Acids Res. 18:6180-6180(1990).
 RP [18]
 RP VARIANT AORTIC ANEURYSM ARG-786.
 RX MEDLINE=91056145; PubMed=2243125;
 RA Kontusaari S., Tromp G., Kuivaniemi H., Romanic A.M., Prockop D.J.;
 RT "A mutation in the gene for type III procollagen (COL3A1) in a family
 RT with aortic aneurysms.";
 RL J. Clin. Invest. 86:1465-1473(1990).
 RP [19]
 RP VARIANT EDS-IV ARG-828.
 RX MEDLINE=94016385; PubMed=8411057;
 RA Richards A.J., Narcisi P., Lloyd J.C., Ferguson C., Pope F.M.;
 RT "The substitution of glycine 661 by arginine in type III collagen
 RT produces mutant molecules with different thermal stabilities and
 RT causes Ehlers-Danlos syndrome type IV.";
 RL J. Med. Genet. 30:690-693(1993).
 RP [20]
 RP VARIANT EDS-IV SER-957.
 RX MEDLINE=89109135; PubMed=2492273;
 RA Tromp G., Kuivaniemi H., Shikata H., Prockop D.J.;
 RT "A single base mutation that substitutes serine for glycine 790 of
 RT the alpha 1 (III) chain of type III procollagen exposes an arginine
 RT and causes Ehlers-Danlos syndrome IV.";
 RL J. Biol. Chem. 264:1349-1352(1989).
 RP [21]
 RP VARIANT EDS-IV VAL-960.


```
CC -----
DR EMBL; AF121217; AAD41775.1; -.
DR InterPro; IPR008161; C1g helix.
DR InterPro; IPR008160; Collagen.
DR InterPro; IPR000885; Fib_collagen_C.
DR Pfam; PF01410; COLFI; 1.
DR Pfam; PF01391; Collagen; 18.
DR ProDom; PD000007; C1g_helix; 4.
DR ProDom; PD003078; Fib_collagen_C; 1.
DR SMART; SM00038; COLFI; 1.
KW Extracellular matrix; Connective tissue; Repeat; Hydroxylation;
KW Glycoprotein; Collagen; Signal; Pyridone carboxylic acid.
FT SIGNAL 1 24
FT PROPEP 25 85
FT CHAIN 86 1108
FT PROPEP 1109 1372
FT SITE 783 785
FT SITE 828 830
FT SITE 1011 1013
FT MOD_RES 86 86
FT MOD_RES 90 90
FT CARBOHYD 1273 1373
FT CONFLICT 132 132
FT CONFLICT 137 137
FT CONFLICT 145 422
FT CONFLICT 431 432
FT CONFLICT 494 494
FT CONFLICT 497 497
FT CONFLICT 502 790
FT CONFLICT 825 825
FT SEQUENCE 1372 AA; 129564 MW; B0693771A8BD20A72 CRC64;

Query Match 64.5%; Score 764; DB 1; Length 1372;
Best Local Similarity 66.8%; Pred. No. 5.8e-30;
Matches 139; Conservative 15; Mismatches 54; Indels 0; Gaps 0;

QY 1 GPPGPGTGLPPGGERGGRGRCFGADGVAGKPGAGRGSPGPGAGKSGPGEAGRP 60
DB 388 GSGFGGAGAGPGLRGSFOSRLPGADRAGVGMPPGNGSTGPGVGPNGDAGR 447

QY 61 GSGALPGAGLTPSGSPGDKTGPAGQDGRPGPPGPGAGQAGVGMGFPKGA 120
DB 448 GSPGLMGRGLPGSPGVNVPAGKEGVLPGIDGRPGIPGAPRGAGNIGFPKGPS 507

QY 121 GPGKAGRGVDPGAGVGPAGKDEAGAGQPPGPGAGRGEGGPGAGSGFGQLGPA 180
DB 508 GDPGRGEGKHGFLAGARGAPDGNNGAQGPFGQVGGKGGQGPAGPFGQGLPGPS 567

QY 181 GPPGAGKPGEGQVGDLCAPGSPGAG 208
DB 568 GTAGEVGKPGERGLPGEGFLPGAPRG 595

RESULT 14
ID CA13 MOUSE
AC P08121; Q61429; O3CRN7;
DT 01-AUG-1988 (Rel. 08, Created)
DT 15-JUL-1999 (Rel. 38, Last sequence update)
DT 15-MAR-2004 (Rel. 43, Last annotation update)
DE Collagen alpha 1(III) chain precursor.
GN COL3A1.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6 X DBA; TISSUE=Embryo;
RX MEDLINE=95011609; PubMed=7926795;
```

```
RA Toman D., de Crombrughe B.;
RT "The mouse type-III procollagen-encoding gene: genomic cloning and
RL complete DNA sequence.";
RN Gene 147:161-168(1994).
RC [2]
RP SEQUENCE FROM N.A.
RX STRAIN=C57BL/6; TISSUE=Brain;
RA MEDLINE=22388257; PubMed=12477932;
RA Klausner R.L., Feingold E.A., Grouse L.H., Derge J.G., Schuler G.D.,
RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schaefer C.F., Bhat N.K.,
RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Hsieh F.,
RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
RA Brownstein M.J., Uedlin T.B., Toshiyuki S., Carninci P., Prange C.,
RA Raha S.S., Loquellano N.A., Peters G.O., Abramson R.D., Mullany S.J.,
RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA Richards S., Worley K.C., Hale S.A., Garcia A.M., Gay L.J., Huiyk S.W.,
RA Villalon D.K., Muzny D.M., Sodergren B.J., Lu X., Gibbs R.A.,
RA Fabey J., Helton E., Kettelman M., Madan A., Rodrigues S., Sanchez A.,
RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,
RA Butterfield Y.S.N., Krzywinski M.I., Skalska U., Smalilus D.E.,
RA Schnerch A., Schein J.E., Jones S.J.M., Marra M.A.;
RT "Generation and initial analysis of more than 15,000 full-length
RL human and mouse cDNA sequences.";
RN Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
RC [3]
RP SEQUENCE OF 1-488 FROM N.A.
RX MEDLINE=88167858; PubMed=3443309;
RA Wood L., Theriault N., Vogeli G.;
RT "Complete nucleotide sequence of the N-terminal domains of the murine
RL alpha-1 type-III collagen chain.";
RN Gene 61:225-230(1987).
RC [4]
RP SEQUENCE OF 1-28 FROM N.A.
RX MEDLINE=85131189; PubMed=3972847;
RA Liaw G., Mudryj M., de Crombrughe B.;
RT "Identification of the promoter and first exon of the mouse alpha 1
RL (III) collagen gene.";
RN J. Biol. Chem. 260:3773-3777(1985).
RC [5]
RP SEQUENCE OF 810-1464 FROM N.A.
RX STRAIN=C57BL/6J; TISSUE=Embryonic head;
RA MEDLINE=21085660; PubMed=11217851;
RA Kawai J., Shinagawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,
RA Arakawa T., Hara A., Fukunishi Y., Konno H., Adachi J., Fukuda S.,
RA Aizawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamataka I.,
RA Saito T., Okazaki Y., Gojobori T., Bono H., Kasukawa T., Saito R.,
RA Kadota K., Matsuda H.A., Ashburner M., Batalov S., Casavant T.,
RA Fleischmann W., Gaasterland T., Gibbs C., King B., Kochiwa H.,
RA Kuehl P., Lewis S., Matsuo Y., Nikaido I., Pesole G., Quackenbush J.,
RA Schriml L.M., Stauble F., Suzuki R., Tomita M., Wagner L., Washio T.,
RA Sakai K., Boffelli D., Bojunga N., Aono H., Baldarelli R., Barsh G.,
RA Brownstein M.J., Bull C., Fletcher C., Fujita M., Gariboldi M.,
RA Gustincich S., Hill D., Hofmann M., Hume D.A., Kamiya M., Lee N.H.,
RA Lyons P., Marchionni L., Mashima J., Mazzarelli J., Mombaerts P.,
RA Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,
RA Saeki H., Sato K., Schoenbach C., Seva T., Shibata Y., Storch K.-F.,
RA Suzuki H., Toyooka K., Wang K.H., Weitz C., Whittaker C., Wilming L.,
RA Wynshaw-Boris A., Yoshida K., Hasegawa Y., Kawaji H., Kohseki S.,
RA Hayashizaki Y.;
RT "Functional annotation of a full-length mouse cDNA collection.";
RN Nature 409:685-690(2001).
RC [6]
RP SEQUENCE OF 1442-1464 FROM N.A.
RX STRAIN=C57BL/6J;
RA MEDLINE=91274355; PubMed=2054384;
RA Metsaeranta M., Toman D., de Crombrughe B., Vuorio E.;
RT "Specific hybridization probes for mouse type I, II, III and IX
RL collagen mRNAs.";
```

RL Biochim. Biophys. Acta 1089:241-243(1991).
 CC !- FUNCTION: Collagen type III occurs in most soft connective tissues
 CC along with type I collagen.
 CC !- SUBUNIT: Trimers of identical alpha 1(III) chains. The chains are
 CC linked to each other by interchain disulfide bonds. Trimers are
 CC also cross-linked via hydroxyllysines.
 CC !- PTM: Proline residues at the third position of the tripeptide
 CC repeating unit (G-X-Y) are hydroxylated in some or all of the
 CC chains.
 CC !- PTM: O-linked glycan consists of a Glc-Gal disaccharide bound to
 CC the oxygen atom of a post-translationally added hydroxyl group (By
 CC similarity).
 CC !- SIMILARITY: Contains 1 WFCC domain.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 CC -----
 CC EMBL; X52046; CAA36279.1; --
 CC EMBL; BC043089; AAH43089.1; --
 CC EMBL; BC058724; AAH58724.1; --
 CC EMBL; M18933; AAA37338.1; --
 CC EMBL; K03037; --; NOT ANNOTATED_CDS.
 CC EMBL; AK019448; BAB31724.1; --
 CC EMBL; X57983; CAA41048.1; --
 CC PIR; A27353; A27353.
 CC PIR; S59856; S59856.
 CC MGI; 88453; Col3a1.
 CC InterPro; IPR008161; Cig_helix.
 CC InterPro; IPR008160; Collagen.
 CC InterPro; IPR000885; Fib_collagen_C.
 CC InterPro; IPR002181; Fibrinogen_C.
 CC InterPro; IPR001007; VWF_C.
 CC Pfam; PF01410; COLFI; 1.
 CC Pfam; PF01391; Collagen; 18.
 CC ProDom; PFD000007; Cig_helix; 1.
 CC ProDom; PFD002078; Fib_collagen_C; 1.
 CC SMART; SMO0038; COLFI; 1.
 CC SMART; SMO0214; VWC; 1.
 CC PROSITE; PS01208; VWF_C; 1.
 CC PROSITE; PS01084; VWF_C; 2; 1.
 CC Extracellular matrix; Connective tissue; Repeat; Hydroxylation;
 KW Glycoprotein; Collagen; Signal.
 FT SIGNAL 1 23
 FT PROPEP 24 154
 FT CHAIN 155 1203
 FT PROPEP 1204 1464
 FT DOMAIN 31 90
 FT DOMAIN 155 169
 FT DOMAIN 170 1195
 FT DOMAIN 1196 1464
 FT CATH 1196 1464
 FT MOD_RES 262 262
 FT MOD_RES 262 262
 FT MOD_RES 283 283
 FT MOD_RES 859 859
 FT MOD_RES 976 976
 FT MOD_RES 1093 1093
 FT MOD_RES 1105 1105
 FT DISULFID 1195 1195
 FT DISULFID 1196 1196
 FT SEQUENCE 1464 AA; 138944 MW; 2104EC27A886090B CRC64;
 Query Match 64.4%; Score 763; DB 1; Length 1464;
 Best Local Similarity 65.4%; Pred. No. 6.7e-30;
 Matches 136; Conservative 13; Mismatches 59; Indels 0; Gaps 0;
 Qy 1 GPPGPGPTGLPFGPGGSGFGGPGGAGVAGPGAGGSGPGAGPGKSGFGGAGRP 60
 Db 467 GSPGPGAGLPGAAGGSGFGGPGGAGVAGPGAGGSGPGAGPGKSGFGGAGRP 526

Qy 61 GEAGLPAGKLTGSPGSGPGDKTGPAGGPGGARGGPGGPGAGVAGVGFPGKGA 120
 Db 527 GTPGPGPIKMGPGSGPGGNDGKPPPGSGSGRPPGPGSGRPPGPGKND 586
 Qy 121 GEPGKAGERGVPPGAVGPGAGKDGAGAGQGGPGGPGAGGEGQPGAGSGFGQLGPGA 180
 Db 587 GAPGXNGERGPGGPGLPGLPGPACKNGETGPGGPGGTGTGAGDKGDSGPPGQLQIPGTG 646
 Qy 181 GPGGAGKPGEGGVPGDILGAPGPGAG 208
 Db 647 GPPGNGKPGEPGPGXGPGVAGPGAGPGKG 674
 RESULT 15
 ID_CAL3_BOVIN STANDARD; PRT; 1049 AA.
 AC P04258;
 DT 20-MAR-1987 (Rel. 04, Created)
 DT 20-MAR-1987 (Rel. 04, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Collagen alpha 1(III) chain.
 GN COL3A1.
 OS Bos taurus (Bovine).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
 OC Bovidae; Bovinae; Bos.
 OC NCBI_TaxID=9913;
 RN [1]
 RP SEQUENCE OF 1-242.
 RX MEDLINE=80026026; PubMed=488906;
 RA Fietzek P.P., Allmann H., Rautenberg J., Henkel W., Wächter E.,
 RA Kuhn K.;
 RT "The covalent structure of calf skin type III collagen. I. The amino
 RT acid sequence of the amino terminal region of the alpha 1(III) chain
 RT (positions 1-222).";
 RL Hoppe-Seyler's Z. Physiol. Chem. 360:809-820(1979).
 RN [2]
 RP SEQUENCE OF 243-422.
 RX MEDLINE=80026027; PubMed=488907;
 RA Dewes H., Fietzek P.P., Kuhn K.;
 RT "The covalent structure of calf skin type III collagen. II. The amino
 RT acid sequence of the cyanogen bromide peptide alpha 1(III)C81,8,10,2
 RT (positions 223-402).";
 RL Hoppe-Seyler's Z. Physiol. Chem. 360:821-832(1979).
 RN [3]
 RP SEQUENCE OF 423-571.
 RX MEDLINE=80026028; PubMed=488908;
 RA Bentz H., Fietzek P.P., Kuhn K.;
 RT "The covalent structure of calf skin type III collagen. III. The
 RT amino acid sequence of the cyanogen bromide peptide alpha 1(III)C84
 RT (positions 403-551).";
 RL Hoppe-Seyler's Z. Physiol. Chem. 360:833-840(1979).
 RN [4]
 RP SEQUENCE OF 572-808.
 RX MEDLINE=80026029; PubMed=488909;
 RA Lang H., Glanville R.W., Fietzek P.P., Kuhn K.;
 RT "The covalent structure of calf skin type III collagen. IV. The amino
 RT acid sequence of the cyanogen bromide peptide alpha 1(III)C85
 RT (positions 552-788).";
 RL Hoppe-Seyler's Z. Physiol. Chem. 360:841-850(1979).
 RN [5]
 RP SEQUENCE OF 809-947.
 RX MEDLINE=80026030; PubMed=488910;
 RA Dewes H., Fietzek P.P., Kuhn K.;
 RT "The covalent structure of calf skin type III collagen. V. The amino
 RT acid sequence of the cyanogen bromide peptide alpha 1(III)C89A
 RT (position 789-927).";
 RL Hoppe-Seyler's Z. Physiol. Chem. 360:851-860(1979).
 RN [6]
 RP SEQUENCE OF 948-1049.
 RX MEDLINE=80026031; PubMed=488911;
 RA Allmann H., Fietzek P.P., Glanville R.W., Kuhn K.;

"The covalent structure of calf skin type III collagen. VI. The amino acid sequence of the carboxyterminal cyanogen bromide peptide alpha 1(III)CB9B (positions 928-1028).";
RL Hope-Seyler's Z. Physiol. Chem. 360:861-868(1979).
CC -!- FUNCTION: Collagen type III occurs in most soft connective tissues along with type I collagen.
CC -!- SUBUNIT: Trimers of identical alpha 1(III) chains. The chains are linked to each other by interchain disulfide bonds. Trimers are also cross-linked via hydroxylsines.
CC -!- FMW: Prolines at the third position of the tripeptide repeating unit (G-X-Y) are hydroxylated in some or all of the chains.
DR PIR; A02862; CGB07S.
DR InterPro; IPR008161; Clg_helix.
DR InterPro; IPR008160; Collagen.
DR InterPro; IPR001007; VWF_C.
DR Pfam; PF01391; Collagen_17.
DR PRODOM; PDOM00007; Clg_helix; 3.
DR PROSITE; PS01208; VWF_C_1; PARTIAL.
KW Extracellular matrix; Connective tissue; Repeat; Hydroxylation;
KW Glycoprotein; Collagen.
FT DOMAIN 1 14 NONHELICAL REGION (N-TERMINAL).
FT DOMAIN 15 1040 TRIPLE-HELICAL REGION.
FT DOMAIN 1041 1049 NONHELICAL REGION (C-TERMINAL).
FT MOD_RES 95 95 HYDROXYLATION.
FT MOD_RES 107 107 HYDROXYLATION.
FT MOD_RES 119 119 HYDROXYLATION.
FT MOD_RES 938 938 HYDROXYLATION.
FT MOD_RES 950 950 HYDROXYLATION.
FT CARBOHYD 107 107 O-LINKED (GAL. : :).
FT CARBOHYD 950 950 O-LINKED (GAL. : :).
FT DISULFID 1040 1040 INTERCHAIN.
FT DISULFID 1041 1041 INTERCHAIN.
SQ SEQUENCE 1049 AA; 93651 MW; 8EEC33D1C66EC9A3 CRC64;

Query Match 64.3%; Score 761.5; DB 1; Length 1049;
Best Local Similarity 60.4%; Pred. No. 6.3e-30;
Matches 139; Conservative 15; Mismatches 55; Indels 21; Gaps 1;
QY 1 GPPGEGFTGLPGPPGSGRSGFPAGDGVAGFKGAPAGRGSPGAPGKSGPGEAGRP 60
DB 312 GSPGEPGANGLPGAAGRGVPGFRGAGNGLPGKGGPPGDRGGPGAPRGVAGEPGRN 371
QY 61 GEAGLPGNKGLTSGSGSPGPKTGPAGQDGRPPGPPGPPGAGQAGVNGFPGKGA 120
DB 372 GLPGGGLRGLPGSGSGKPGKPGSGQGETGRGPPGSPGPRGQGVNGFPGKGN 431
QY 121 GEPGKAGRGVPGPPGAVGPGAGKDGAGAGQAGPPGAPGAGRGQAGSPGFGQLPGPA 180
DB 432 GAPKNGRGVGGGPGPGPGAGKNGETGPGQPPGTGPGSGDKGDTGPPGQGLGLEGTS 491
QY 181 GPPGKAGVPGGQGVFDLGP-----GPSGPAGG 209
DB 492 GPPGNGKPGGPGKGEAGAFPGKGDGAPGERGPPGAGGPPGPRGG 541

Search completed: September 24, 2004, 11:07:23
Job time : 7.27357 secs

This Page Blank (uspto)

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: September 24, 2004, 11:06:56 ; Search time 8.04914 Seconds
(without alignments)
2497.663 Million cell updates/sec

Title: US-10-658-989a-1

Perfect score: 1184

Sequence: 1 GPPPEPQTGLPQPPGERG.....GEGVFGDLGAPGSPAGG 209

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 283366 seqs, 96191526 residues

Total number of hits satisfying chosen parameters: 283366

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : PIR 78:**

1: pir1.*

2: pir2.*

3: pir3.*

4: pir4.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1149	97.0	1464	1 CGHU1S	collagen alpha 1(I)
2	1118	94.4	671	1 CGRTUS	collagen alpha 1(I)
3	1116	94.3	1453	2 S21626	collagen alpha 1(I)
4	1067	90.1	1042	1 CGCH1S	collagen alpha 1(I)
5	862	72.8	1487	1 CGHU6C	collagen alpha 1(I)
6	860	72.6	1418	2 T45467	collagen alpha 1(I)
7	850	71.8	1486	1 B40333	collagen alpha 1(I)
8	847	71.5	1492	2 A40333	collagen alpha 1(I)
9	839	70.9	1419	2 A41182	collagen alpha 1(I)
10	839	70.9	1487	2 B41182	collagen alpha 1(I)
11	824	69.6	779	1 CGBO1S	collagen alpha 1(I)
12	815	68.8	673	1 CGBO6C	collagen alpha 1(I)
13	794.5	67.1	886	2 I50694	collagen alpha 1(I)
14	786	66.4	1496	1 CGHU2V	collagen alpha 2(V)
15	776	65.5	1466	1 CGHU7L	collagen alpha 1(I)
16	765	64.6	1497	2 I43607	procollagen type V
17	763	64.4	1464	2 S59856	collagen alpha 1(I)
18	761.5	64.3	1049	1 CGBO7S	collagen alpha 1(I)
19	760	64.2	1373	1 A43291	collagen alpha 2(I)
20	744	62.8	1366	1 CGHU2S	collagen alpha 2(I)
21	629.5	53.2	1027	2 S28774	collagen alpha 1(V)
22	607	51.3	1838	1 CGHU1V	collagen alpha 1(I)
23	602.5	50.9	310	2 I50696	collagen alpha 2(I)
24	600.5	50.7	1414	1 S23809	collagen alpha 1(I)
25	599	50.6	1843	2 S18803	collagen alpha 1(V)
26	590.5	49.9	632	2 S42731	collagen alpha 1(C)
27	589	49.7	964	1 CGCH2S	collagen alpha 2(I)
28	577	48.7	1806	1 CGHU1E	collagen alpha 1(X)
29	576.5	48.7	730	2 A36226	collagen alpha 1 C

RESULT 1

CGHU1S

collagen alpha 1(I) chain precursor - human

N;Alternate names: procollagen alpha 1(I) chain

C;Species: Homo sapiens (man)

C;Date: 12-Aug-1981 #sequence revision 04-Oct-1996 #text_change 31-Dec-2000

C;Accession: I60114; S01143; A93335; I55254; A39943; I55237; S09400; B90567; S11-

5289; A29439; I53466; A02852; I37247

R;D'Alessio, M.; Bernard, M.; Pretorius, P.J.; de Wet, W.; Ramirez, F.; Pretorius, P.J.

Gene 67, 105-115, 1988

A;Title: Complete nucleotide sequence of the region encompassing the first twenty-five ex

A;Reference number: I60114; MUID:89329734; PMID:2843432

A;Accession: I60114

A;Status: translated from GB/EMBL/DBJ

A;Molecule type: DNA

A;Residues: 1-369, 1-371-589 <DAL>

A;Cross-references: GB:M20789; NID:g179593; PIDN:AAB59373.1; PID:g179594

R;Tromp, G.; Kuvaniemi, H.; Stacey, A.; Shikata, H.; Baldwin, C.T.; Jaenisch, R.; Prock-

Biochem. J. 253, 919-922, 1988

A;Title: Structure of a full-length cDNA clone for the prepro-alpha-1(I) chain of human t

A;Reference number: S01143; MUID:89025644; PMID:3178743

A;Accession: S01143

A;Molecule type: mRNA

A;Residues: 1-472 <TRO>

A;Cross-references: EMBL:X07884; NID:g30015; PIDN:CAA30731.1; PID:g30016; GB:M36546; NID

R;Chu, M.L.; de Wet, W.; Bernard, M.; Ding, J.F.; Morabito, M.; Myers, J.; Williams, C.;

Nature 310, 337-340, 1984

A;Title: Human proalpha1(I) collagen gene structure reveals evolutionary conservation of

A;Reference number: A93335; MUID:84270697; PMID:6462220

A;Accession: A93335

A;Molecule type: DNA

A;Residues: 1-58, 1-60-181 <CHU>

A;Cross-references: EMBL:X00820; NID:g35657; PIDN:CAA25394.1; PID:g35658

R;Rossouw, C.M.S.; Vergeer, W.P.; du Plooy, S.J.; Bernard, M.P.; Ramirez, F.; de Wet, W.;

J. Biol. Chem. 262, 15151-15157, 1987

A;Title: DNA sequences in the first intron of the human pro-alpha 1(I) collagen gene enha

A;Reference number: I55254; MUID:88033098; PMID:2822714

A;Accession: I55254

A;Status: translation not shown; translated from GB/EMBL/DBJ

A;Molecule type: DNA

A;Residues: 1-45 <ROS>

A;Cross-references: GB:J02829; NID:g180387; PIDN:AAA51993.1; PID:g180388

R;Bornstein, P.; McKay, J.; Morishima, J.K.; Devarayalu, S.; Gelinas, R.E.

Proc. Natl. Acad. Sci. U.S.A. 84, 8869-8873, 1987

A;Title: Regulatory elements in the first intron contribute to transcriptional control of

A;Reference number: A39943; MUID:98097389; PMID:3480516

A;Accession: A39943

A;Molecule type: DNA

A;Residues: 1-34 <BOR>

A;Cross-references: GB:J03559; NID:g180876; PIDN:AAA52052.1; PID:g553238

R;Chu, M.L.; de Wet, W.; Bernard, M.; Ramirez, F.

J. Biol. Chem. 260, 2315-2320, 1985
A>Title: Fine structural analysis of the human pro-alpha 1 (I) collagen gene. Promoter s
A:Reference number: I55237; MUID:85130970; PMID:2857713
A:Accession: I55237
A:Status: translation not shown; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-34 <CH2>
A:Cross-references: GB:M10627; NID:gl180383; PIDN:AAAS1992.1; PID:G553226
R:Wittz, M.K.; Keene, D.R.; Hori, H.; Glanville, R.W.; Steinmann, B.; Rao, V.H.; Hollist
J. Biol. Chem. 265, 6312-6317, 1990
A>Title: In vivo and in vitro noncovalent association of excised alpha1(I) amino-terminal
rome, type VII.
A:Reference number: A35233; MUID:90202908; PMID:2318855
A:Accession: A35233
A:Molecule type: protein
A:Residues: 33-52 <WR>
A>Note: this propeptide fragment remained non-covalently bound to a defective, uncleaved
R:Well, D.; d'Alessio, M.; Ramirez, F.; de Wet, W.; Cole, W.G.; Chan, D.; Bateman, J.F.
EMBO J. 8, 1705-1710, 1989
A>Title: A base substitution in the exon of a collagen gene causes alternative splicing
A:Reference number: S09400; MUID:89356643; PMID:2767050
A:Accession: S09400
A:Molecule type: mRNA
A:Residues: 156-183 <WEI>
R:Click, E.M.; Bornstein, P.
Biochemistry 9, 4699-4706, 1970
A>Title: Isolation and characterization of the cyanogen bromide peptides from the alpha
A:Reference number: A90567; MUID:71038625; PMID:5529814
A:Contents: CNBr0-1, CNBr2, CNBr4, CNBr5
A:Accession: B90567
A:Molecule type: protein
A:Residues: 169-198, 'Z', 200-201, 'Z', 203-206, 'Z', 208-209, 'Z', 211-228, 'B', 230, 'BB', 233, 'Z'
A:Experimental source: skin
A>Note: evidence for 170-allysine
R:Baeger, B.; Notbohm, H.; Diebold, J.; Lehmann, H.; Bodo, M.; Deutzmann, R.; Mueller, P.
Eur. J. Biochem. 192, 153-159, 1990
A>Title: A critical crosslink region in human-bone-derived collagen type I. Specific cle
A:Reference number: S11372; MUID:90382436; PMID:2169412
A:Accession: S11372
A:Molecule type: protein
A:Residues: 175-187, 274-287, 'P', 289 <BAE>
A>Note: sequence of collagen alpha 1(S)(I) isolated from bone after pepsin digestion
R:Deak, S.B.; Scholz, P.M.; Amenta, P.S.; Constantinou, C.D.; Levi-Minzi, S.A.; Gonzalez
J. Biol. Chem. 266, 21827-21832, 1991
A>Title: The substitution of arginine for glycine 85 of the alpha 1(I) procollagen chain
operatively melting of intact type I collagen.
A:Reference number: I55342; MUID:92042092; PMID:1718984
A:Accession: I55342
A:Status: translated from GB/EMBL/DBJ
A:Molecule type: mRNA
A:Residues: 258-268, 1347-1357 <DEA>
A:Cross-references: GB:S67495; NID:G239007; PIDN:AAAB20350.1; PID:G239008
A>Note: sequences from the 5' and 3' ends only are shown; mutant sequence 263-Arg report
R:Morgan, P.H.; Jacobs, H.G.; Segrest, J.P.; Cunningham, L.W.
J. Biol. Chem. 245, 5042-5048, 1970
A>Title: Comparative study of glycopeptides derived from selected vertebrate collagens.
A:Reference number: A92069; MUID:71001508; PMID:4319110
A:Accession: A92069
A:Molecule type: protein
A:Residues: 263-268 <MOR>
A:Experimental source: skin
A>Note: attachment of 2-O-alpha-D-glucosyl-O-beta-D-galactose to 5-hydroxylysine
R:Labhard, M.E.; Hollister, D.W.
Matrix 10, 124-130, 1990
A>Title: Segmental amplification of the entire helical and telopeptide regions of the ch
A:Reference number: S15989; MUID:90326017; PMID:2374517
A:Accession: S15989
A:Molecule type: mRNA
A:Residues: 281-302, 402-420, 823-843, 925-944, 1026-1045, 1143-1162 <LAB>
R:Wittz, M.K.; Rao, V.H.; Glanville, R.W.; Labhard, M.E.; Pretorius, P.J.; de Vries, W.N.
Connect. Tissue Res. 29, 1-11, 1993
A>Title: A cysteine for glycine substitution at position 175 in an alpha 1 (I) chain of
A:Reference number: I52905; MUID:93339042; PMID:8339541
A:Accession: I52905

A:Accession: I52905
A:Status: translated from GB/EMBL/DBJ
A:Molecule type: mRNA
A:Residues: 342-352, 'C', 354-359 <W12>
A:Cross-references: GB:S64717; NID:G408195; PIDN:AAAB27677.1; PID:G408196
A>Note: mutant sequence from patient with osteogenesis imperfecta
R:Bernard, M.P.; Chu, M.L.; Myers, J.C.; Ramirez, F.; Eikenberry, E.F.; Prockop, D.J.
Biochemistry 22, 5213-5223, 1983
A>Title: Nucleotide sequences of complementary deoxyribonucleic acids for the proalpha 1
A:Reference number: A90476; MUID:84080385; PMID:6689127
A:Accession: A90476
A:Molecule type: mRNA
A:Residues: 425-1250, 'X', 1252-1328, 'S', 1330-1390, 'X', 1392-1464 <BER>
A:Cross-references: GB:K01228; NID:G180391; PIDN:AAAS1955.1; PID:G180392
A>Note: sequence partially completed for missing nucleotides by A29439
R:Chu, M.L.; Gargiulo, V.; Williams, C.J.; Ramirez, F.
J. Biol. Chem. 260, 691-694, 1985
A>Title: Multixon deletion in an osteogenesis imperfecta variant with increased type III
A:Reference number: A22161; MUID:85104934; PMID:2981843
A:Accession: A22161
A:Molecule type: DNA
A:Residues: 472-594, 'R', 596-607 <CH3>
A:Cross-references: GB:K03178; GB:K03179; NID:G179612; NID:G179613; PIDN:AAAS1847.1; PID:
A>Note: the authors translated the codon CGT for residue 595 as Pro
R:Wallis, G.A.; Starman, B.J.; Zinn, A.B.; Byers, P.H.
Am. J. Hum. Genet. 46, 1034-1040, 1990
A>Title: Variable expression of osteogenesis imperfecta in a nuclear family is explained
A:Reference number: A35336; MUID:90252792; PMID:2339700
A:Accession: A35336
A:Molecule type: mRNA
A:Residues: 710-720, 'E', 722-737, 'E', 739-745 <WAL>
A>Note: the authors translated the codons CAG for 721 and CGT for 738 as Glu
R:Forlino, A.; Zolezzi, F.; Valli, M.; Pignatti, P.F.; Cetta, G.; Brunelli, P.C.; Mottes,
Hum. Mol. Genet. 3, 2201-2206, 1994
A>Title: Severe (type III) osteogenesis imperfecta due to glycine substitutions in the c
A:Reference number: I54365; MUID:95187161; PMID:7881420
A:Accession: I54365
A:Status: translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 746-766, 'S', 768-781 <FOR>
A:Cross-references: GB:L47667; NID:G1009093; PIDN:AAAS95576.1; PID:G1009094
R:Chessler, S.D.; Wallis, G.A.; Byers, P.H.
J. Biol. Chem. 268, 18218-18225, 1993
A>Title: Mutations in the carboxyl-terminal propeptide of the pro alpha 1(I) chain of ty
A:Reference number: A47426; MUID:93352646; PMID:8349697
A:Accession: A47426
A:Molecule type: mRNA
A:Residues: 1179-1276, 'H', 1278-1336, 1339-1387, 'R', 1389-1464 <CHE>
A:Cross-references: GB:S64596; NID:G407589; PIDN:AAAB27856.1; PID:G407590
A>Note: sequence extracted from NCBI backbone (NCBIN:136444, NCBI:P:136445)
A>Note: does not represent an experimentally determined sequence but three different muta
A:Accession: B47426
A:Molecule type: mRNA
A:Residues: 1179-1464 <CH4>
A:Experimental source: normal dermal fibroblast culture
A:Accession: C47426
A:Molecule type: mRNA
A:Residues: 1179-1276, 'H', 1278-1464 <CH5>
A:Experimental source: fetal cell 86-237
A:Accession: D47426
A:Molecule type: mRNA
A:Residues: 1179-1336, 1339-1464 <CH6>
A:Experimental source: fetal cell 86-146
A:Accession: E47426
A:Molecule type: mRNA
A:Residues: 1179-1387, 'R', 1389-1464 <CH7>
A:Experimental source: fetal cell 88-251
R:Cohn, D.H.; Apone, S.; Eyre, D.R.; Starman, B.J.; Andreassen, P.; Charbonneau, H.; Nicl
J. Biol. Chem. 263, 14605-14607, 1988
A>Title: Substitution of Cysteine for Glycine within the Carboxyl-terminal Telopeptide of
A:Reference number: I55269; MUID:89008319; PMID:3170557
A:Accession: I55269

A;Experimental source: skin
R;Butler, W.T.
Biochemistry 9, 44-50, 1970
A;Title: Chemical studies on the cyanogen bromide peptides of rat skin collagen. The cova
A;Reference number: A90566; MUID:70085124; PMID:5411206
A;Contents: CNBr5
A;Accession: A90566
A;Molecule type: protein
A;Residues: 103-139 <BU2>
A;Experimental source: skin
R;Balian, G.; Click, E.M.; Bornstein, P.
Biochemistry 10, 4470-4478, 1971
A;Title: Structure of rat skin collagen alpha1-CB8. Amino acid sequence of the hydroxylar
A;Reference number: A90357; MUID:72136131; PMID:4335087
A;Contents: CNBr8
A;Accession: A90357
A;Molecule type: protein
A;Residues: 140-238 <BA1>
A;Experimental source: skin
R;Balian, G.; Click, E.M.; Hermodson, M.A.; Bornstein, P.
Biochemistry 11, 3798-3806, 1972
A;Title: Structure of rat skin collagen alpha1-CB8. Amino acid sequence of the hydroxylar
A;Reference number: A90362; MUID:73008942; PMID:4342027
A;Contents: CNBr8
A;Accession: A90362
A;Molecule type: protein
A;Residues: 239-418 <BA2>
A;Experimental source: skin
R;Butler, W.T.; Underwood, S.P.; Finch Jr., J.E.
Biochemistry 13, 2945-2953, 1974
A;Title: Chemical studies on the cyanogen bromide peptides of rat skin collagen. Amino ac
A;Reference number: A90379; MUID:74271984; PMID:4366532
A;Contents: CNBr3
A;Accession: A90379
A;Molecule type: protein
A;Residues: 419-567 <BU3>
A;Experimental source: skin
R;Stoltz, M.; Timpl, R.; Furthmayr, H.; Kuehn, K.
Eur. J. Biochem. 37, 287-294, 1973
A;Title: Structural and immunogenic properties of a major antigenic determinant in neutr
A;Reference number: A91209; MUID:74011954; PMID:4126850
A;Contents: CNBr6
A;Accession: A91209
A;Molecule type: protein
A;Residues: 568-651 <ST1>
A;Experimental source: skin
A;Note: this region probably corresponds to positions 949-1032 of the alpha 1(I) chain
A;Note: the major antigenic determinant (of neutral salt-extracted rat skin collagen) inv
R;Stoltz, M.; Timpl, R.; Kuehn, K.
FEBS Lett. 26, 61-65, 1972
A;Title: Non-helical regions in rat collagen alpha1-chain.
A;Reference number: A91385; MUID:73049495; PMID:4636751
A;Contents: CNBr6
A;Accession: A91385
A;Molecule type: protein
A;Residues: 651-671 <ST2>
A;Experimental source: skin
A;Note: the composition of peptides comprising residues 1-9 and 1-19 confirms the sequen
A;Note: this region (residues 651-671 above) probably corresponds to positions 1032-1052
C;Comment: Prolines and lysines at the third position of the tripeptide repeating unit (o
ed and subsequently O-glycosylated.
C;Comment: The order of the nine CNBr peptides in the alpha 1(I) chain of rat skin colla
C;Comment: The complete chain contains 1052 residues.
C;Superfamily: collagen alpha 1(I) chain; fibrillar collagen carboxyl-terminal homology;
C;Keywords: blocked amino end; coiled coil; extracellular matrix; glycoprotein; hydroxyl
F;1/Modified site: blocked amino end (Glx) (probably pyrrolidone carboxylic acid) #statu
F;9/Modified site: allysine (lys) #status experimental
F;103,424,547/Binding site: carbohydrate (lys) (covalent) #status experimental
F;103/Modified site: 5-hydroxylysine (lys) #status experimental
F;424,547/Modified site: 5-hydroxylysine (lys) (partial) #status experimental

Query Match 94.4%; Score 1118; DB 1; Length 671;
Best Local Similarity 90.0%; Pred. No. 1.9e-55;
Query Local

Query Match 90.1%; Score 1067; DB 1; Length 1042;
 Best Local Similarity 86.4%; Pred. No. 1.8e-52;
 Matches 190; Conservative 4; Mismatches 14; Indels 12; Gaps 1;

QY 1 GPP-----GPPGPTGLDPPGGRGPGSRFFPGNDGVAGKGPAGRGSPGPA 48
 DB 296 GPPGPAEGEKRGGPAGLPGPAGRGAPSRFFPGADGIAGPKPPGGRGSPGAV 355
 QY 49 GPKSPGAGRPAGLPGKGLTGSPGSPDCKTTPPGAGQDGRPGPPGPPGARGQA 108
 DB 356 GPKSPGAGRPAGLPGKGLTGSPGSPDCKTTPPGAGQDGRPGPPGPPGARGQA 415
 QY 109 GVMGPPGPKGAAGBFGKAGRGVPGPCGAVGPAKQGEAGAGQGGPPGAGRGEGGCPA 168
 DB 416 GVMGPPGPKGAAGBFGKAGRGVPGPCGAVGPAKQGEAGAGQGGPPGAGRGEGGCPA 475
 QY 169 GSPGFGGLPGPAGPPGAGKPGEGGVGCDLCAPGSPGAG 208
 DB 476 GAGFGGLPGPAGPPGAGKPGEGGVGCDLCAPGSPGAG 515

RESULT 5
 CCHU6C
 collagen alpha 1(II) chain precursor [validated] - human
 N/Alternate names: procollagen alpha 1(II) chain
 N/Contains: chondrocalcin; collagen alpha 1(II) chain precursor splice form 1; collagen
 C/Species: Homo sapiens (man)
 C/Date: 28-May-1986 #sequence revision 01-Sep-1995 #text change 08-Dec-2000
 A/Accession: A38513; S06715; S24270; A24828; S06496; A35428; A30147; A33116; S64674; S63750; I37251; I37252; I37253; I37254; I55338; I59535; I61910
 R/Ryan, M.C.; Sieraski, M.; Sandell, L.J.
 Genomics 8, 41-48, 1990
 A/Title: The human type II procollagen gene: identification of an additional protein-cod
 A/Reference number: A38513; MUID:91184811; PMID:2081599
 A/Accession: A38513
 A/Molecule type: DNA
 A/Residues: 1-103 <RYA>
 A/Cross-references: GB:M60299; NID:g180883; PIDN:AAA73873.1; PID:g180884
 R/Su, M.W.; Lee, B.; Ramirez, F.; Machado, M.; Horton, W.
 Nucleic Acids Res. 17, 9473, 1989
 A/Title: Nucleotide sequence of the full length cDNA encoding for human type II procolla
 A/Reference number: S06715; MUID:90067946; PMID:2587267
 A/Accession: S06715
 A/Molecule type: mRNA
 A/Residues: 1-28, 'R', 99-1487 <SU2>
 A/Cross-references: EMBL:X16468; NID:g29515; PIDN:CAA34488.1; PID:g29516
 A/Note: alternative splice form 1
 R/Vikkula, M.; Metsaeranta, M.; Syvaenen, A.C.; Ala-Kokko, L.; Vuorio, E.; Peltonen, L.
 Biochem. J. 285, 287-294, 1992
 A/Title: Structural analysis of the regulatory elements of the type-II procollagen gene.
 A/Reference number: S24270; MUID:92344585; PMID:1637314
 A/Accession: S24270
 A/Status: translation not shown
 A/Molecule type: DNA
 A/Residues: 1-28 <VIK>
 A/Cross-references: EMBL:X58709; GB:S40537; NID:g35659
 A/Note: This translation is not annotated in GenBank entry HSPROCOE1, release 111.0
 R/Nunez, A.M.; Kohno, K.; Martin, G.R.; Yamada, Y.
 Gene 44, 11-16, 1986
 A/Title: Promoter region of the human pro-alpha-1(II)-collagen gene.
 A/Reference number: A24828; MUID:87031574; PMID:3021582
 A/Accession: A24828
 A/Molecule type: DNA
 A/Residues: 1-8, 'T', 10-28 <NUN>
 A/Cross-references: GB:M25698; NID:g180872; PIDN:AAA52051.1; PID:g553237
 R/Baldwin, C.T.; Reginator, A.M.; Smith, C.; Jimenez, S.A.; Prockop, D.J.
 Biochem. J. 262, 521-528, 1989
 A/Title: Structure of cDNA clones coding for human type II procollagen. The alpha-1(II)
 A/Reference number: S06496; MUID:90026318; PMID:2803268
 A/Accession: S06496
 A/Molecule type: mRNA
 A/Residues: 7-28, 'R', 99-157, 'P', 159-440, 'G', 442-456, 'E', 458-640, 'A', 642-831, 'PA', 834, 'F',

A/Cross-references: EMBL:X16711; NID:g30040; PIDN:CAA34683.1; PID:g30041
 A/Note: alternative splice form 1
 R/Ryan, M.C.; Sandell, L.J.
 J. Biol. Chem. 265, 10334-10339, 1990
 A/Title: Differential expression of a cysteine-rich domain in the amino-terminal propept
 A/Reference number: A35428; MUID:90285153; PMID:2355003
 A/Accession: A35428
 A/Status: not compared with conceptual translation
 A/Molecule type: mRNA
 A/Residues: 27-81, 'L', 83-103 <RYA2>
 A/Note: alternative splice form 2; splicing appears to be under developmental regulation
 R/Su, M.W.; Benson-Chanda, V.; Vissing, H.; Ramirez, F.
 Genomics 4, 438-441, 1989
 A/Title: Organization of the exons coding for Pro alpha-1(II) collagen N-propeptide conf
 A/Reference number: A30147; MUID:89233138; PMID:2714801
 A/Accession: A30147
 A/Molecule type: DNA
 A/Residues: 104-157, 'P', 159-236 <SUM>
 A/Cross-references: GB:J03065; GB:M23660; GB:M25655; GB:M25730; GB:M32168; GB
 R/Ala-Kokko, L.; Baldwin, C.T.; Moskowitz, R.W.; Prockop, D.J.
 Proc. Natl. Acad. Sci. U.S.A. 87, 6565-6568, 1990
 A/Title: Single base mutation in the type II procollagen gene (COL2A1) as a cause of pri
 A/Reference number: A94227; MUID:90370826; PMID:1975693
 A/Accession: A33116
 A/Molecule type: DNA
 A/Residues: 171-172, 'C', 174-175 <ALA>
 A/Note: mutant sequence from a family with primary generalized osteoarthritis
 R/Diab, M.; Wu, J.J.; Eyre, D.R.
 Biochem. J. 314, 327-332, 1996
 A/Title: Collagen type IX from human cartilage: a structural profile of intermolecular c
 A/Reference number: S64673; MUID:96195147; PMID:8660302
 A/Accession: S64674
 A/Molecule type: protein
 A/Residues: 188-189, 'X', 191-195; 1224-1230, 'X', 1232-1236 <DIA>
 R/Franc, S.; Marzin, E.; Boutillon, M.M.; Lafont, R.; Lechene de la Porte, P.; Herbage,
 Eur. J. Biochem. 234, 125-131, 1995
 A/Title: Immunohistochemical and biochemical analyses of 20000-25000-year-old fossil ear
 A/Reference number: S63514; MUID:96096730; PMID:8529631
 A/Accession: S63514
 A/Molecule type: protein
 A/Residues: 243-261, 575-590; 756-763, 'X', 765-779 <FRA>
 R/Tiller, G.E.; Weiss, M.A.; Polombo, P.A.; Gruber, H.E.; Rimoin, D.L.; Cohn, D.H.; Eyre,
 Am. J. Hum. Genet. 56, 388-395, 1995
 A/Title: An RNA-splicing mutation (G-SIVS20) in the type II collagen gene (COL2A1) in a
 A/Reference number: I38867; MUID:95150028; PMID:7847372
 A/Accession: I38867
 A/Status: preliminary; translated from GB/EMBL/DDBJ
 A/Molecule type: DNA
 A/Residues: 440, 'G', 442-456, 'E', 458-480, 'P', 482-509 <TIL1>
 A/Cross-references: EMBL:U15195; NID:g557053; PIDN:AA660370.1; PID:g557054
 R/Ramirez, F.
 submitted to the EMBL Data Library, December 1988
 A/Reference number: S04892
 A/Accession: S04892
 A/Molecule type: mRNA
 A/Residues: 501-676, 'A', 678-783, 'A', 785-831, 'PA', 834, 'F', 836-1214 <RAM>
 A/Cross-references: EMBL:X13783; NID:g30037; PIDN:CAA32030.1; PID:g930050
 R/Vikkula, M.; Peltonen, L.
 FEBS Lett. 250, 171-174, 1989
 A/Title: Structural analyses of the polymorphic area in type II collagen gene.
 A/Reference number: S05000; MUID:89325561; PMID:2753125
 A/Accession: S05000
 A/Molecule type: DNA
 A/Residues: 630-640, 'A', 642-785 <VIK2>
 A/Cross-references: EMBL:X16158; NID:g29951; PIDN:CAA34278.1; PID:g1335018; PIDN:CAA3427
 PIDN:CAA3428.1; PID:g1335023; PIDN:CAA34284.1; PID:g1335024
 R/Bogaert, R.; Tiller, G.E.; Weiss, M.A.; Gruber, H.E.; Rimoin, D.L.; Cohn, D.H.; Eyre, D
 J. Biol. Chem. 267, 22522-22526, 1992
 A/Title: An amino acid substitution (Gly853--Sclu) in the collagen alpha 1(II) chain pro
 A/Reference number: A44309; MUID:93054548; PMID:1429602
 A/Accession: A44309
 A/Status: nucleic acid sequence not shown; not compared with conceptual translation
 A/Molecule type: DNA; mRNA

A:Residues: 752-831, 'FA', 834, 'F', 836-1005, 'K', 1007-1036, 'Q', 1038-1052, 'E', 1054-1068, 'T', 'A';
A:Cross-references: GB:L00977; NID:g180812; PIDN:AA23914.1; PID:g258774
A:Note: sequence extracted from NCBI backbone (NCBIP:117273); parts of this sequence were
A:Note: this translation is not annotated and this publication is not cited in GenBank
A:Note: mutant sequence associated with perinatal lethal hypochondrogenesis
R:Tiller, G.E.; Rimo, D.L.; Murray, L.W.; Cohn, D.H.
Proc. Natl. Acad. Sci. U.S.A. 87, 3889-3893, 1990
A:Title: Tandem duplication within a type II collagen gene (COL2A1) exon in an individual
A:Reference number: S16502; MUID:90251662; PMID:2339128
A:Accession: S16502
A:Molecule type: DNA
A:Residues: 1164-1184, 'GSPGKGANGIPGPI', 1185-1199, 'TIL2'
A:Cross-references: ENBL:M37126; NID:g180808; PIDN:AA52037.1; PID:g180809
A:Note: mutant sequence from a patient with spondyloepiphyseal dysplasia
R:Cheah, K.S.E.; Stoker, N.G.; Griffin, J.R.; Grosfeld, F.G.; Solomon, E.
Proc. Natl. Acad. Sci. U.S.A. 82, 2555-2559, 1985
A:Title: Identification and characterization of the human type II collagen gene (COL2A1)
A:Reference number: A02858; MUID:85190534; PMID:3857598
A:Accession: A02858
A:Molecule type: DNA
A:Residues: 1032-1056, 'N', 1058-1068, 'T', 1070-1487, 'CHE'
A:Cross-references: GB:J00116; NID:g180395; PIDN:AA51997.1; PID:g180396
R:Elima, K.; Vuorio, T.; Vuorio, E.
Nucleic Acids Res. 15, 9499-9504, 1987
A:Title: Determination of the single polyadenylation site of the human pro-alpha-1(II) chain
A:Reference number: A27280; MUID:88067771; PMID:2825137
A:Accession: A27280
A:Molecule type: mRNA
A:Residues: 1175-1487, 'ELI'
A:Cross-references: ENBL:X06268; NID:g30096; PIDN:CAA29604.1; PID:g30097
A:Experimental source: fetal epiphyseal cartilage
R:van der Rest, M.; Rosenberg, L.C.; Olsen, B.R.; Poole, A.R.
Biochem. J. 237, 923-925, 1986
A:Title: Chondrocalcin is identical with the C-propeptide of type II procollagen.
A:Reference number: A57033; MUID:87099927; PMID:3800925
A:Accession: A57033
A:Molecule type: protein
A:Residues: 'X', 1244-1246, 'N', 1248, 'X', 1250-1265; 1295-1305; 1395-1408, 'VAN'
A:Note: chondrocalcin identified as released collagen 1(II) chain carboxyl-terminal propeptide
R:Strom, C.M.; Upholt, W.B.
Nucleic Acids Res. 12, 1025-1038, 1984
A:Title: Isolation and characterization of genomic clones corresponding to the human type II collagen gene
A:Reference number: A21733; MUID:84118798; PMID:6320112
A:Accession: A21733
A:Molecule type: DNA
A:Residues: 1245-1295, 'STR1'
A:Cross-references: ENBL:X00339; ENBL:X00298; NID:g394699; PIDN:CAA25092.1; PID:g4378975
A:Accession: B21733
A:Molecule type: DNA
A:Residues: 894-909, 'PE', 'STR2'
A:Cross-references: GB:K01785; NID:g30035; PIDN:CAA25082.1; PID:g13335032
R:Nunez, A.M.; Francomano, C.; Young, M.F.; Martin, G.R.; Yamada, Y.
Biochemistry 24, 6343-6348, 1985
A:Title: Isolation and partial characterization of genomic clones coding for a human pro alpha 1(II) chain
A:Reference number: A24561; MUID:86104139; PMID:3002437
A:Accession: A24561
A:Molecule type: DNA
A:Residues: 1296-1358, 'NUN2'
A:Cross-references: GB:M12048; NID:g180017
A:Note: this translation is not annotated in GenBank entry HUMCCT2A, release 111.0
R:Sangiorji, F.O.; Benson-Chanda, V.; de Wet, W.J.; Sobel, M.E.; Tsipouras, P.; Ramirez, N.
Nucleic Acids Res. 13, 2207-2225, 1985
A:Title: Isolation and partial characterization of the entire human pro alpha 1(II) chain
A:Reference number: I37249; MUID:85215609; PMID:2987845
A:Accession: S59491
A:Molecule type: DNA
A:Residues: 7-28, 'R', 99-114; 541-578; 786-802; 1055-1056, 'N', 1058-1068, 'T', 1070-1109; 1200-1201
A:Accession: 184453
A:Status: translated from GB/ENBL/DBU
A:Molecule type: DNA
A:Residues: 7-28, 'SAN2'

A:Cross-references: GB:M23759; NID:g180845; ENBL:X03320; GB:M24938; NID:g30104
A:Note: the GenBank PID is based on an incorrect reading frame
A:Accession: I37250
A:Status: translated from GB/ENBL/DBU
A:Molecule type: DNA
A:Residues: 541-560, 'SAN3'
A:Cross-references: ENBL:X02378; GB:M23870; NID:g30107; PIDN:CAA26227.1; PID:g929621
A:Accession: I37251
Query Match 72.8%; Score 862; DB 1; Length 1487;
Best Local Similarity 74.0%; Pred. No. 5,1e-41;
Matches 154; Conservative 11; Mismatches 44; Indels 0; Gaps 0;
QY 1 GPPGPGTGLPGLPGRGGSPGPGADGVAGPKGPGAGRGQGGPAGSGPFGQLPGPA 60
Db 492 GARGPGGVGPIGPPGGRGAPGNRFPQDGLAGKAGPGRGSGLAGKAGNGDPGR 551
QY 61 GEAGLPKAGLGTSGSPGPGKTPGPGAGDGRPPGPGGPGGARGOAGVGMFFGPKGAA 120
Db 552 GEPGLPGARGLTGRFDAGPQGVGPGGAPGEGDGRPPGPGGARGOAGVGMFFGPKGAN 611
QY 121 GEPGKAGRGVPPGAVGPGKAGCAQAGPAGPAGRGQGGPAGSGPFGQLPGPA 180
Db 612 GEPGKAGRGVPPGAVGPGKAGCAQAGPAGPAGRGQGGPAGSGPFGQLPGPA 671
QY 181 GPPGAGRGVPPGAVGPGKAGCAQAGPAGPAGRGQGGPAGSGPFGQLPGPA 208
Db 672 GPPGAGRGVPPGAVGPGKAGCAQAGPAGPAGRGQGGPAGSGPFGQLPGPA 699
RESULT 6
T45467
collagen alpha 1(II) chain precursor [imported] - horse
N:Alternate names: type II collagen
C:Species: Equus caballus (domestic horse)
C:Date: 31-Jan-2000 #sequence_revision 31-Jan-2000 #text_change 04-Mar-2000
C:Accession: T45467
R:Richardson, D.W.; Dodge, G.R.
Submitted to the ENBL Data Library, June 1996
A:Description: Cloning of equine type II collagen and modulation of its expression in eq
A:Reference number: Z22977
A:Accession: T45467
A:Status: preliminary; translated from GB/ENBL/DBU
A:Molecule type: mRNA
A:Residues: 1-1418, 'RIC'
A:Cross-references: ENBL:U62528; PIDN:AA805773.1
C:Superfamily: collagen alpha 1(II) chain; fibrillar collagen carboxyl-terminal homology;
Query Match 72.6%; Score 860; DB 2; Length 1418;
Best Local Similarity 74.0%; Pred. No. 6.4e-41;
Matches 154; Conservative 10; Mismatches 44; Indels 0; Gaps 0;
QY 1 GPPGPGTGLPGLPGRGGSPGPGADGVAGPKGPGAGRGQGGPAGSGPFGQLPGPA 60
Db 423 GARGPGGVGPIGPPGGRGAPGNRFPQDGLAGKAGPGRGSGLAGKAGNGDPGR 482
QY 61 GEAGLPKAGLGTSGSPGPGKTPGPGAGDGRPPGPGGPGGARGOAGVGMFFGPKGAA 120
Db 483 GEPGLPGARGLTGRFDAGPQGVGPGGAPGEGDGRPPGPGGARGOAGVGMFFGPKGAN 542
QY 121 GEPGKAGRGVPPGAVGPGKAGCAQAGPAGPAGRGQGGPAGSGPFGQLPGPA 180
Db 543 GEPGKAGRGVPPGAVGPGKAGCAQAGPAGPAGRGQGGPAGSGPFGQLPGPA 602
QY 181 GPPGAGRGVPPGAVGPGKAGCAQAGPAGPAGRGQGGPAGSGPFGQLPGPA 208
Db 603 GPPGAGRGVPPGAVGPGKAGCAQAGPAGPAGRGQGGPAGSGPFGQLPGPA 630
RESULT 7
B40333
collagen alpha 1(II) chain precursor - African clawed frog
C:Species: Xenopus laevis (African clawed frog)

C;Date: 10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change 10-Sep-1999
 C;Accession: B40333
 R;Su, M.W.; Suzuki, H.R.; Bleker, J.J.; Solursh, M.; Ramirez, F.
 J. Cell Biol. 115, 565-575, 1991
 A;Title: Expression of two nonallelic type II procollagen genes during Xenopus laevis em
 F;37-96/Domain: von Willebrand factor type C repeat homology <VMC>
 A;Reference number: A40333; MUID:92011898; PMID:1918153
 A;Accession: B40333
 A;Status: preliminary
 A;Molecule type: mRNA
 A;Residues: 1-1486 <SUA>
 A;Cross-references: GB:M63595
 C;Superfamily: collagen alpha 1(I) chain; fibrillar collagen carboxyl-terminal homology;
 C;Keywords: coiled coil; extracellular matrix; glycoprotein; trimer; triple helix
 F;1258-1486/Domain: fibrillar collagen carboxyl-terminal homology <FCC>
 Query Match 71.8%; Score 850; DB 1; Length 1486;
 Best Local Similarity 73.1%; Pred. No. 2.4e-40;
 Matches 152; Conservative 11; Mismatches 45; Indels 0; Gaps 0;
 QY 1 GPPGEPPTGLPPGRRGSGRFPAGADGAGVAGPKGAPGAGSGPPGAGKSGPGEAGRP 60
 DB 494 GARGEPGAAGPAGPGRGAFGNRFPQDGLAGPKGAPGARGVPLGLGKGGNGDFGRP 553
 QY 61 GEAGLPKAGLGTGSPGSPGDKTGPAGADGCRPPGPPGARGGAGVWGFPGPKGAA 120
 DB 554 GEPGLPARGLTGRPDAGPQGVKVPAGSGEDGRPPFPQAGRGQGVWGFPGPKGAN 613
 QY 121 GEPKAGRGVPPGPPGAVGAGKDGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 180
 DB 614 GEPKAGKGLVAGLPLGKDGRTGSGQNPAGPAGRGAGAGAGAGAGAGAGAGAG 673
 QY 181 GPPGKAGRGVPPGPPGAGVPPGAGVPPGAGVPPGAGVPPGAGVPPGAGVPP 208
 DB 674 GSPGEGKPGDQGVFGAGAPGLVGRG 701
 RESULT 8
 A40333
 collagen alpha 1'(II) chain precursor - African clawed frog
 C;Species: Xenopus laevis (African clawed frog)
 C;Date: 16-Sep-1992 #sequence_revision 16-Sep-1992 #text_change 16-Jul-1999
 C;Accession: A40333
 R;Su, M.W.; Suzuki, H.R.; Bleker, J.J.; Solursh, M.; Ramirez, F.
 J. Cell Biol. 115, 565-575, 1991
 A;Title: Expression of two nonallelic type II procollagen genes during Xenopus laevis em
 F;37-96/Domain: von Willebrand factor type C repeat homology <VMC>
 A;Reference number: A40333; MUID:92011898; PMID:1918153
 A;Accession: A40333
 A;Status: nucleic acid sequence not shown
 A;Molecule type: mRNA
 A;Residues: 1-1492 <SUA>
 A;Cross-references: GB:M63596
 A;Note: this sequence is presented as substitutions relative to another sequence in a fi
 es they replace; the appropriate interpretation of the sequence figure was reconstructed
 C;Superfamily: collagen alpha 1(I) chain; fibrillar collagen carboxyl-terminal homology;
 C;Keywords: coiled coil; extracellular matrix; glycoprotein; trimer; triple helix
 F;1263-1492/Domain: fibrillar collagen carboxyl-terminal homology <FCC>
 Query Match 71.5%; Score 847; DB 2; Length 1492;
 Best Local Similarity 73.6%; Pred. No. 3.5e-40;
 Matches 153; Conservative 9; Mismatches 46; Indels 0; Gaps 0;
 QY 1 GPPGEPPTGLPPGRRGSGRFPAGADGAGVAGPKGAPGAGSGPPGAGKSGPGEAGRP 60
 DB 497 GARGEPGAAGPAGPGRGAFGNRFPQDGLAGPKGAPGARGVPLGLGKGGNGDFGRP 556
 QY 61 GEAGLPKAGLGTGSPGSPGDKTGPAGADGCRPPGPPGARGGAGVWGFPGPKGAA 120
 DB 557 GEPGLPARGLTGRPDAGPQGVKVPAGSGAEDGRPPFPQAGRGQGVWGFPGPKGAN 616
 QY 121 GEPKAGRGVPPGPPGAVGAGKDGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 180

617 GPPGKAGKGLGGAPGLRGLPKDGTGAQGNPAGPAGRGEGQPPGSPGFGQLPGPP 676
 QY 181 GPPGKAGKPGEGQVPGDGLGAPGSPGAG 208
 DB 677 GSPGEGKPGDQGVFGAGAGLVGRG 704
 RESULT 9
 A41182
 collagen alpha 1(II) chain precursor - mouse
 C;Species: Mus musculus (house mouse)
 C;Date: 28-May-1992 #sequence_revision 28-May-1992 #text_change 13-Aug-1999
 C;Accession: A41182; A44885
 R;Metsaeranta, M.; Toman, D.; de Crombrughe, B.; Vuorio, E.
 J. Biol. Chem. 266, 16862-16869, 1991
 A;Title: Mouse type II collagen gene. Complete nucleotide sequence, exon structure, and
 A;Reference number: A41182; MUID:91358489; PMID:1885613
 A;Accession: A41182
 A;Status: preliminary; not compared with conceptual translation
 A;Molecule type: DNA
 A;Residues: 1-1419 <MET>
 A;Cross-references: GB:M65161
 R;Cheah, K.S.; Lau, E.T.; Au, P.K.; Tam, P.P.
 Development 111, 945-953, 1991
 A;Title: Expression of the mouse alpha 1(II) collagen gene is not restricted to cartilag
 A;Reference number: A44885; MUID:91347939; PMID:1879363
 A;Accession: A44885
 A;Molecule type: DNA
 A;Residues: 1-28 <CHE>
 A;Cross-references: GB:S63190; NID:9234368; PIDN:AAB19627.1; PID:9234369
 A;Note: sequence extracted from NCBI backbone (NCBI:63190, NCBI:63192)
 C;Superfamily: collagen alpha 1(I) chain; fibrillar collagen carboxyl-terminal homology;
 C;Keywords: alternative splicing; coiled coil; extracellular matrix; glycoprotein; trime
 F;1191-1419/Domain: fibrillar collagen carboxyl-terminal homology <FCC>
 Query Match 70.9%; Score 839; DB 2; Length 1419;
 Best Local Similarity 72.1%; Pred. No. 9.3e-40;
 Matches 150; Conservative 12; Mismatches 46; Indels 0; Gaps 0;
 QY 1 GPPGEPPTGLPPGRRGSGRFPAGADGAGVAGPKGAPGAGSGPPGAGKSGPGEAGRP 60
 DB 424 GARGPFGGAGTGPGRGAFGNRFPQDGLAGPKGAPGARGVPLGLGKGGNGDFGRP 483
 QY 61 GEAGLPKAGLGTGSPGSPGDKTGPAGADGCRPPGPPGARGGAGVWGFPGPKGAA 120
 DB 484 GEPGLPARGLTGRPDAGPQGVKVPAGSGEDGRPPFPQAGRGQGVWGFPGPKGAN 543
 QY 121 GEPKAGRGVPPGPPGAVGAGKDGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 180
 DB 544 GEPGKAGRGVPPGPPGAVGAGKDGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 603
 QY 181 GPPGKAGKPGEGQVPGDGLGAPGSPGAG 208
 DB 604 GPPGEGKQDQGIPEAGAGPLVGRG 631
 RESULT 10
 B41182
 collagen alpha 1(II) chain precursor (long splice form) - mouse
 C;Species: Mus musculus (house mouse)
 C;Date: 28-May-1992 #sequence_revision 28-May-1992 #text_change 16-Jul-1999
 C;Accession: B41182
 R;Metsaeranta, M.; Toman, D.; de Crombrughe, B.; Vuorio, E.
 J. Biol. Chem. 266, 16862-16869, 1991
 A;Title: Mouse type II collagen gene. Complete nucleotide sequence, exon structure, and
 A;Reference number: A41182; MUID:91358489; PMID:1885613
 A;Accession: B41182
 A;Status: preliminary; not compared with conceptual translation
 A;Molecule type: DNA
 A;Residues: 1-1487 <NET>
 A;Cross-references: GB:M65161
 C;Superfamily: collagen alpha 1(I) chain; fibrillar collagen carboxyl-terminal homology;
 C;Keywords: alternative splicing; coiled coil; extracellular matrix; glycoprotein; trime

A;Accession: A91201
 A;Molecule type: protein
 A;Residues: 295-562 <FI3>
 A;Experimental source: skin
 A;Wendt, P.; Mark, K.V.D.; Rexrodt, F.; Kuehn, K.
 Eur. J. Biochem. 30, 169-183, 1972
 A;Title: The covalent structure of collagen. The amino-acid sequence of the 112 residues
 A;Reference number: A91201; PMID:73042276; PMID:4343808
 A;Accession: A91201
 A;Molecule type: protein
 A;Residues: 563-675 <WEN>
 A;Experimental source: skin
 A;Fietzek, P.P.; Rexrodt, F.W.; Wendt, P.; Stark, M.; Kuehn, K.

R; Butler, W.T.; Finch Jr., J.E.; Miller, E.J.

J. Biol. Chem. 252, 639-643, 1977
A:Title: The covalent structure of cartilage collagen. Evidence for sequence heterogeneity
A:Reference number: A92210; MUID:7703864; PMID:833147
A:Accession: A92210
A:Molecule type: protein
A:Residues: 139-178, 'Z', 180-184, 'PA', 187-190, 'AS', 193-194, 'T', 196-198 <BU2>
A:Cross-references: EMBL:U07973; NID:9520454; PIDN:AAA83407.1; PID:G537432
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: mRNA
A:Residues: 1-886 <NAH>
A:Cross-references: EMBL:U07973; NID:9520454; PIDN:AAA83407.1; PID:G537432
C:Genetics:
A:Gene: COL1A1
C:Superfamily: collagen alpha 1(I) chain; fibrillar collagen carboxyl-terminal homology;
F:30-50/Domain: von Willebrand factor type C repeat homology <WVC>
Query Match 67.1%; Score 794.5; DB 2; Length 886;
Best Local Similarity 65.9%; Pred. No. 1.9e-37;
Matches 143; Conservative 13; Mismatches 55; Indels 9; Gaps 1;
QY 1 GPP-----GEPGPTGLPGPPGGERGCGSRGPGADGVAGVAGPKGAPGAGRGSPGAPGPK 51
DB 458 GPPGEGKRGANGEPQGVPTGCRGSPGRLGSGNGLPGKGPAGRGSPGPPGSP 517
QY 52 GSPGAGRPGAGLPGAKLGTGSPGPDGKTGTPGAGQDGRPGPPGARGAQGV 111
DB 518 GPAGRGDGGGGLPGMLPGIIPGSDGPKGPPGNGQGPGRSGPPGAPGPPGQGV 577
QY 112 GPPGKGAAGEPGKAGRGVPGPGVAGVAGPKGAGAGAGQGPAGRGAGRGSPG 171
DB 578 GPPGKGNAGAGKNGRGPGGPGTGPAGKNGDVGLPGPPGAPGAGDRGEPGSPG 637
QY 172 GFQGLPGPAGPGHAGKPGGQGVDPDLGAPGSPGAG 208
DB 638 GLQGLPGPGPAGENGKPGGPKDGGGPPGPKG 674
RESULT 14
CGH2V
collagen alpha 2(V) chain precursor - human
C:Species: Homo sapiens (man)
C:Date: 31-Jul-1989 #sequence revision 28-Jul-1995 #text change 31-Dec-2000
C:Accession: A31427; A54555; S43643; A25874; I55239; I59025; A25374; A30017
R:Woodbury, D.; Benson-Chanda, V.; Ramirez, F.
J. Biol. Chem. 264, 2735-2738, 1989
A:Title: Amino-terminal propeptide of human pro-alpha2(V) collagen conforms to the structure of alpha2(V) chain precursor
A:Reference number: A31427; MUID:89123368; PMID:2914927
A:Accession: A31427
A:Molecule type: mRNA
A:Residues: 1-463 <WOO>
A:Cross-references: GB:J04478; NID:G179697; PIDN:AAA51859.1; PID:G179698
A:Experimental source: placenta
R:Greenspan, D.S.; Lee, S.T.; Lee, B.S.; Hoffman, G.G.
Gene Expr. 1, 29-39, 1991
A:Title: Homology between alpha2(V) and alpha1(III) collagen promoters and evidence for a common regulatory region
A:Reference number: A54555; MUID:92314691; PMID:1820205
A:Accession: A54555
A:Molecule type: DNA
A:Residues: 1-32 <GRE>
A:Cross-references: GB:M58529; NID:G180834; PIDN:AA41699.1; PID:G553235
R:Moradi-Ameili, M.; Rousseau, J.C.; Kleman, J.P.; Champilaud, M.F.; Boutillon, M.M.; Berr
Eur. J. Biochem. 221, 987-995, 1994
A:Title: Diversity in the processing events at the N-terminus of type-V collagen.
A:Reference number: S43642; MUID:94237164; PMID:8181482
A:Accession: S43642
A:Molecule type: protein
A:Residues: 288-291, 'P', 293-294, 'X', 296-297, 606, 'X', 608-617 <MOR>
R:Well, D.; Bernard, M.; Gargano, S.; Ramirez, F.
Nucleic Acids Res. 15, 181-198, 1987
A:Title: The pro alpha 2(V) collagen gene is evolutionarily related to the major fibrillar
A:Reference number: A25874; MUID:87146331; PMID:3029669
A:Accession: A25874
A:Molecule type: mRNA; DNA
A:Residues: 398-1496 <WRI>
A:Cross-references: GB:X04758; NID:G29588; PIDN:CAA28454.1; PID:G1340175
A:Experimental source: rhabdomyosarcoma cell line
R:Myers, J.C.; Loidl, H.R.; Stolle, C.A.; Seyer, J.M.
J. Biol. Chem. 260, 5533-5541, 1985

J. Biol. Chem. 252, 639-643, 1977
A:Title: The covalent structure of cartilage collagen. Evidence for sequence heterogeneity
A:Reference number: A92210; MUID:7703864; PMID:833147
A:Accession: A92210
A:Molecule type: protein
A:Residues: 139-178, 'Z', 180-184, 'PA', 187-190, 'AS', 193-194, 'T', 196-198 <BU2>
A:Cross-references: EMBL:U07973; NID:9520454; PIDN:AAA83407.1; PID:G537432
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: mRNA
A:Residues: 1-886 <NAH>
A:Cross-references: EMBL:U07973; NID:9520454; PIDN:AAA83407.1; PID:G537432
C:Genetics:
A:Gene: COL1A1
C:Superfamily: collagen alpha 1(I) chain; fibrillar collagen carboxyl-terminal homology;
F:30-50/Domain: von Willebrand factor type C repeat homology <WVC>
Query Match 67.1%; Score 794.5; DB 2; Length 886;
Best Local Similarity 65.9%; Pred. No. 1.9e-37;
Matches 143; Conservative 13; Mismatches 55; Indels 9; Gaps 1;
QY 1 GPP-----GEPGPTGLPGPPGGERGCGSRGPGADGVAGVAGPKGAPGAGRGSPGAPGPK 51
DB 458 GPPGEGKRGANGEPQGVPTGCRGSPGRLGSGNGLPGKGPAGRGSPGPPGSP 517
QY 52 GSPGAGRPGAGLPGAKLGTGSPGPDGKTGTPGAGQDGRPGPPGARGAQGV 111
DB 518 GPAGRGDGGGGLPGMLPGIIPGSDGPKGPPGNGQGPGRSGPPGAPGPPGQGV 577
QY 112 GPPGKGAAGEPGKAGRGVPGPGVAGVAGPKGAGAGAGQGPAGRGAGRGSPG 171
DB 578 GPPGKGNAGAGKNGRGPGGPGTGPAGKNGDVGLPGPPGAPGAGDRGEPGSPG 637
QY 172 GFQGLPGPAGPGHAGKPGGQGVDPDLGAPGSPGAG 208
DB 638 GLQGLPGPGPAGENGKPGGPKDGGGPPGPKG 674
RESULT 14
CGH2V
collagen alpha 2(V) chain precursor - human
C:Species: Homo sapiens (man)
C:Date: 31-Jul-1989 #sequence revision 28-Jul-1995 #text change 31-Dec-2000
C:Accession: A31427; A54555; S43643; A25874; I55239; I59025; A25374; A30017
R:Woodbury, D.; Benson-Chanda, V.; Ramirez, F.
J. Biol. Chem. 264, 2735-2738, 1989
A:Title: Amino-terminal propeptide of human pro-alpha2(V) collagen conforms to the structure of alpha2(V) chain precursor
A:Reference number: A31427; MUID:89123368; PMID:2914927
A:Accession: A31427
A:Molecule type: mRNA
A:Residues: 1-463 <WOO>
A:Cross-references: GB:J04478; NID:G179697; PIDN:AAA51859.1; PID:G179698
A:Experimental source: placenta
R:Greenspan, D.S.; Lee, S.T.; Lee, B.S.; Hoffman, G.G.
Gene Expr. 1, 29-39, 1991
A:Title: Homology between alpha2(V) and alpha1(III) collagen promoters and evidence for a common regulatory region
A:Reference number: A54555; MUID:92314691; PMID:1820205
A:Accession: A54555
A:Molecule type: DNA
A:Residues: 1-32 <GRE>
A:Cross-references: GB:M58529; NID:G180834; PIDN:AA41699.1; PID:G553235
R:Moradi-Ameili, M.; Rousseau, J.C.; Kleman, J.P.; Champilaud, M.F.; Boutillon, M.M.; Berr
Eur. J. Biochem. 221, 987-995, 1994
A:Title: Diversity in the processing events at the N-terminus of type-V collagen.
A:Reference number: S43642; MUID:94237164; PMID:8181482
A:Accession: S43642
A:Molecule type: protein
A:Residues: 288-291, 'P', 293-294, 'X', 296-297, 606, 'X', 608-617 <MOR>
R:Well, D.; Bernard, M.; Gargano, S.; Ramirez, F.
Nucleic Acids Res. 15, 181-198, 1987
A:Title: The pro alpha 2(V) collagen gene is evolutionarily related to the major fibrillar
A:Reference number: A25874; MUID:87146331; PMID:3029669
A:Accession: A25874
A:Molecule type: mRNA; DNA
A:Residues: 398-1496 <WRI>
A:Cross-references: GB:X04758; NID:G29588; PIDN:CAA28454.1; PID:G1340175
A:Experimental source: rhabdomyosarcoma cell line
R:Myers, J.C.; Loidl, H.R.; Stolle, C.A.; Seyer, J.M.
J. Biol. Chem. 260, 5533-5541, 1985

F.201/Modified site: allylsine (Lys) #status predicted
F.290, 293, 296, 608, 614, 1004, 1007, 1013, 1028, 1034/Modified site: 4-hydroxyproline (Pro) #status predicted
F.299, 1139/Modified site: 5-hydroxylysine (Lys) #status predicted
F.299, 1139/binding site: carboxylate (Lys) (covalent) #status predicted
F.1035/Modified site: 5-hydroxylysine (Lys) #status experimental
F.1250-1251/Cleavage site: Glu-Asp (procollagen C-endopeptidase) #status predicted
F.1259, 1397/Binding site: carboxylate (Asn) (covalent) #status predicted
F.1233, 1299, 1325/Disulfide bonds: interchain #status predicted
F.1333-1494, 1402-1447/Disulfide bonds: #status predicted

```

Query Match      66.4%; Score 786; DB 1; Length 1496;
Best Local Similarity 67.3%; Pred. No. 8.3e-37;
Matches 140; Conservative 15; Mismatches 53; Indels 0; Gaps 0;

1  GPPGPGGTGLPPGPPGGRGGPSRGGFCADGAVGPKPACGERSGPGAGPKXGSPGEACRP 60
501 GPRDPTGLTLPFGVPGERGAPENKRGFGSDGLPFPKAQERGVPVSGSGPKXGSGDGP 560

61 GEAGLPAAKGLTSGPSGPDGKTGPPCPAQDGRPGPPCPGARGAGVGMFGPPKGAA 120
561 GEPGLPARGLTGPNVQGGPEGKLGPLGAPGEDRPGPPSGIGIKQPGTMTGLPXPXSN 620

121 GEPCKACGERGVPPGPGAVGAGKDGAGACAGPPGAPGACGERGQGPAGSGPGGLGPPA 180
621 GDPKPKPEAGNPGVPGQRGAPGKDKGVTPGPPGDLRGERGEGQPPGPTGFGHFGPP 680

181 GPPGACKPGEGQVPGDLAGAPSGGPAG 208

681 GPPGEGGKPDQGVPGGPGAVGLGPRG 708

```

RESULT 15
 CGH07L
 collagen alpha 1(III) chain precursor - human
 Alternate names: procollagen alpha 1(III) chain
 Species: Homo sapiens (man)
 Date: 24-Apr-1984 #sequence revision 01-Sep-1995 #text change 21-Jul-2000
 CAC: Accession: S05272; S04642; PE0011; S01726; S04887; A90399; A94562; I51868; S59511; A904
 Submitted to the EMBL Data Library, February 1989
 Reference number: S05272
 Accession: S05272
 Status: preliminary
 Molecule type: mRNA
 Residues: 1-1240, 'V' 1242-1466 <PRO>
 Cross-references: EMBL:X14420; NID:G30057; PIDN:CAA32583.1; PID:G30058
 Ala-Kokko, L.; Kontusaari, S.; Balgwin, C.T.; Kuivaniemi, H.; Prockop, D.J.
 Biochemistry J. 260, 509-516, 1989
 Title: Structure of cDNA clones coding for the entire prepro-alpha(III) chain of human
 procollagen.
 Reference number: S04642; MUID:89350838; PMID:2764896
 Accession: S04642
 Molecule type: mRNA
 Residues: 1-1196 <ALA>
 Cross-references: EMBL:X14420; NID:G30057; PIDN:CAA32583.1; PID:G30058
 Note: the complete sequence is not shown
 Benson-Chanda, V.; Su, M.W.; Weil, D.; Chu, M.L.; Ramirez, F.
 Gene 78, 255-265, 1989
 Title: Cloning and analysis of the 5' portion of the human type-III procollagen gene (C
 Reference number: PE0011, MUID:89378752; PMID:2777083
 Accession: PE0011
 Molecule type: DNA
 Residues: 1-176 <BEN>
 Cross-references: GB:M26939; NID:G180813; PIDN:AAA52040.1; PID:G180814
 Toman, P.D.; Ricca, G.A.; de Crombrughe, B.
 Nucleic Acids Res. 16, 7201, 1988
 Title: Nucleotide sequence of a cDNA coding for the amino-terminal region of human pre
 Reference number: S01726; MUID:88303360; PMID:3405773
 Accession: S01726
 Molecule type: mRNA
 Residues: 1-170 <TOM>
 Cross-references: EMBL:X07240; NID:G30060; PIDN:CAA30229.1; PID:G30061
 Note: the authors translated the codon CAG for residue 154 as His

R;Janeczko, R.A.; Ramirez, F.
Nucleic Acids Res. 17, 6742, 1989
A;Title: Nucleotide and amino acid sequences of the entire human alpha-1 (III) collagen.
A;Reference number: S04887; MUID:89386015; PMID:2780304
A;Accession: S04887
A;Molecule type: mRNA
A;Residues: 148-163, 'G', 164-240, 'D', 242-471, 'D', 473-487, 'L', 489, 'S', 491-613, 'Y', 615-634,
A;Cross-references: EMBL:X15333; NID:929545; PIDN:CAA33387.1; PID:930045
A;Note: the authors' translation of residues 908-932 is inconsistent with the nucleotide
R;Seyer, J.M.; Kang, A.H.
Biochemistry 16, 1158-1164, 1977
A;Title: Covalent structure of collagen: amino acid sequence of cyanogen bromide peptide
A;Reference number: A90399; MUID:77134724; PMID:557335
A;Accession: A90399
A;Molecule type: protein
A;Residues: 'V', 165-225, 229-232, 'P', 234-292, 'D', 294-398 <SEY1>
A;Experimental source: liver
A;Note: sequence corrected by A94562; attachment of 2-O-alpha-D-glucosyl-O-beta-D-galact
R;Seyer, J.M.
submitted to the Atlas, December 1977
A;Reference number: A94562
A;Accession: A94562
A;Molecule type: protein
A;Residues: 'V', 169-225, 229-277, 'A', 279-292, 'D', 294, 'S', 296-398 <SEY2>
A;Experimental source: liver
A;Note: author submitted corrections to A90399
R;Milewicz, D.M.; Witz, A.M.; Smith, A.C.; Manchester, D.K.; Waldstein, G.; Byers, P.H.
Am. J. Hum. Genet. 53, 62-70, 1993
A;Title: Parental somatic and germ-line mosaicism for a multilexon deletion with unusual
ispring.
A;Reference number: I51868; MUID:93304430; PMID:8317500
A;Accession: I51868
A;Status: preliminary; translated from GB/EMBL/DBJ
A;Molecule type: DNA
A;Residues: 186-194 <ML>
A;Cross-references: GB:S62925; NID:9386425; PIDN:AD1937.1; PID:94261637
R;Chiodo, A.A.; Sillence, D.O.; Cole, W.G.; Bateman, J.F.
Biochem. J. 311, 939-943, 1995
A;Title: Abnormal type III collagen produced by an exon-17-skipping mutation of the COL3
A;Reference number: S59511; MUID:96067614; PMID:7487954
A;Accession: S59511
A;Molecule type: mRNA
A;Residues: 302-423 <CHI>
A;Cross-references: GB:S79877; NID:gl195576; PIDN:AAB35615.1; PID:gl195577
R;Seyer, J.M.; Kang, A.H.
Biochemistry 17, 3404-3411, 1978
A;Title: Covalent structure of collagen: amino acid sequence of five consecutive CNBr pe
A;Reference number: A90414; MUID:79000343; PMID:687591
A;Accession: A90414
A;Molecule type: protein
A;Residues: 399-675, 'N', 677-727 <SEY3>
A;Experimental source: liver
R;Lee, B.; Vitale, E.; Superti-Furga, A.; Steinmann, B.; Ramirez, F.
J. Biol. Chem. 266, 5256-5259, 1991
A;Title: G to T transversion at position +5 of a splice donor site causes skipping of th
A;Reference number: I55349; MUID:91161621; PMID:1672129
A;Accession: I55349
A;Status: translated from GB/EMBL/DBJ
A;Molecule type: DNA
A;Residues: 537-605 <LEE>
A;Cross-references: GB:M59312; NID:gl80815; PIDN:AAA52041.1; PID:gl180816
R;Seyer, J.M.; Mainardi, C.; Kang, A.H.
Biochemistry 19, 1583-1589, 1980
A;Title: Covalent structure of collagen: amino acid sequence of alpha1 (III)-CB5 from ty
A;Reference number: A90438; MUID:80198282; PMID:6246925
A;Accession: A90438
A;Molecule type: protein
A;Residues: 728-895, 'A', 897-964 <SEY4>
A;Experimental source: liver
R;Cole, W.G.; Chiodo, A.A.; Lamande, S.R.; Janeczko, R.; Ramirez, F.; Dahl, H.H.M.; Chan
J. Biol. Chem. 265, 17070-17077, 1990
A;Title: A base substitution at a splice site in the COL3A1 gene causes exon skipping an
A;Reference number: A38303; MUID:91009133; PMID:2145268

A;Accession: A38303
A;Molecule type: mRNA
A;Residues: 861-1015 <COL>
A;Cross-references: GB:J05617; GB:M55603; GB:M59227; NID:gl80878; PIDN:AAB59383.1; PID:gl
A;Note: a mutant sequence with 942-977 spliced out from a patient with Ehlers-Danlos synd
R;Mankoo, B.S.; Daigleish, R.
Nucleic Acids Res. 16, 2337, 1988
A;Title: Human pro alpha1(III) collagen: cDNA sequence for the 3' end.
A;Reference number: S02119; MUID:88189827; PMID:3357782
A;Accession: S02119
A;Status: translation not shown
A;Molecule type: mRNA
A;Residues: 950-1018, 'Y', 1020-1183, 'S', 1185-1466 <MAN>
A;Cross-references: EMBL:X06700; NID:930053; PIDN:CAA29886.1; PID:930054
R;Seyer, J.M.; Kang, A.H.
Biochemistry 20, 2621-2627, 1981
A;Title: Covalent structure of collagen: amino acid sequence of alpha1 (III)-CB9 from tyr
A;Reference number: A90446; MUID:81208139; PMID:7016180
A;Accession: A90446
A;Molecule type: protein
A;Residues: 965-979, 'A', 981-984, 'PS', 987, 'QN', 990-1096, 'P', 1098-1152, 'AT', 1155, 'S', 1157-
A;Experimental source: liver
R;Joldi, H.R.; Brinker, J.M.; May, M.; Pihlajaniemi, T.; Morrow, S.; Rosenbloom, J.; Mye
Nucleic Acids Res. 12, 9383-9394, 1984
A;Title: Molecular cloning and carboxyl-propeptide analysis of human type III procollagen
A;Reference number: A93551; MUID:85087944; PMID:6096827
A;Accession: A93551
A;Molecule type: mRNA
A;Residues: 1065-1155, 'P', 1157-1466 <LOI>
A;Cross-references: EMBL:X01655; EMBL:X01742; NID:929584; PIDN:CAA25821.1
R;Wiskuln, M.; Daigleish, R.; Klueve-Becker, B.; Kennard, S.I.; Tolstoshev, P.; Brant
Biochemistry 25, 1408-1413, 1986
A;Title: Human type III collagen gene expression is coordinately modulated with the type
A;Reference number: I52393; MUID:86187804; PMID:3754462
A;Accession: I52393
A;Status: translated from GB/EMBL/DBJ
A;Molecule type: mRNA
A;Residues: 1161-1200 <M18>
A;Cross-references: GB:M13146; NID:gl80415; PIDN:AAA52003.1; PID:gl80416
R;Emanuel, B.S.; Cannizzaro, L.A.; Seyer, J.M.; Myers, J.C.
Proc. Natl. Acad. Sci. U.S.A. 82, 3385-3389, 1985
A;Title: Human alpha 1(III) and alpha 2(V) procollagen genes are located on the long arm
A;Reference number: I59025; MUID:85216505; PMID:3858826
A;Accession: I59025
A;Status: translated from GB/EMBL/DBJ
A;Molecule type: mRNA
A;Residues: 1165-1196 <EMA>
A;Cross-references: GB:M11134; NID:gl80417; PIDN:AAA52004.1; PID:gl80418
R;Chu, M.L.; Weil, D.; de Wet, W.; Bernard, M.; Sippola, M.; Ramirez, F.
J. Biol. Chem. 260, 4357-4363, 1985
A;Title: Isolation of cDNA and genomic clones encoding human pro-alpha(III) collagen. Pe
A;Reference number: A92516; MUID:85157600; PMID:2579949
A;Accession: A92516
A;Molecule type: DNA
A;Residues: 1176-1240, 'V', 1242-1356, 'P', 1358-1466 <CHU>
A;Cross-references: GB:M10615; GB:M10793; GB:M10795; GB:M10796; GB:M10797; GB
A;Experimental source: liver
A;Note: the authors translated the codon TTC for residue 1057 as Tyr; the codons given fo
ation
C;Comment: Prolines and lysines at the third position of the tripeptide repeating unit (C
3-hydroxylated. About 15% of the lysines are 5-hydroxylated and some are subsequently O
C;Genetics.
A;Gene: GDB:COL3A1
A;Cross-references: GDB:118729; OMIM:120180
A;Map position: 2q31-2q31
A;Introns: 27/1; 94/3; 111/3; 149/3; 176/3; 554/3; 587/3; 1175/3; 1275/1; 1337/3; 1418/3
A;Note: the list of introns is incomplete; defects in this gene can result in Ehlers-Danl
C;Complex: type III collagen is a homotrimer of monomers initially linked by disulfide bo
er of their length, is formed with desmosine cross-links made from lysine and allylsine re
C;Function:
A;Description: structural component of extracellular fibrous polymer that maintains intes
C;Superfamily: collagen alpha 1(I) chain; fibrillar collagen carboxyl-terminal homology;
C;Keywords: coiled coil; Ehlers-Danlos syndrome; extracellular matrix; Glycoprotein; Hyd

Query Match 97.0%; Score 1149; DB 3; Length 1341;
Best Local Similarity 93.6%; Pred. No. 2e-72;
Matches 206; Conservative 0; Mismatches 2; Indels 12; Gaps 1;

QY 1 GPP-----GEGPTGLPGPPGGRGPGSGRPGADGVAGPKGPAGRGSPGPA 48
DB 334 GPPGPAEGEGKRGARGEPTGLPGPPGGRGPGSGRPGADGVAGPKGPAGRGSPGPA 393
QY 49 GPKSGPGEAGRPFGEAGLFGAKGLTGSPGSPGPDGKTGPPGAGQDGRPPGPPGARGQA 108
DB 394 GPKSGPGEAGRPFGEAGLFGAKGLTGSPGSPGPDGKTGPPGAGQDGRPPGPPGARGQA 453
QY 109 GVMGPPGPKGAAGRPFGEAGLFGAKGLTGSPGSPGPDGKTGPPGAGQDGRPPGPPGARGQA 168
DB 454 GVMGPPGPKGAAGRPFGEAGLFGAKGLTGSPGSPGPDGKTGPPGAGQDGRPPGPPGARGQA 513
QY 169 GSPFGQGLPGPAGPPGEAGKPGEGQVPGDLAGPSPGARG 208
DB 514 GSPFGQGLPGPAGPPGEAGKPGEGQVPGDLAGPSPGARG 553

RESULT 2

US-09-500-811-18
; Sequence 18, Application US/09500811
; Patent No. 6323314

GENERAL INFORMATION:
; APPLICANT: Qvist, Per

APPLICANT: Bonde, Martin
; TITLE OF INVENTION: A Method for Assaying Collagen Fragments

; TITLE OF INVENTION: in Body Fluids, A Test Kit and Means for Carrying Out the
; TITLE OF INVENTION: Method and Use of the Method to Diagnose the Presence of
; TITLE OF INVENTION: Disorders Associated with the Metabolism of
; NUMBER OF SEQUENCES: 21

CORRESPONDENCE ADDRESS:
ADDRESSEE: Darby & Darby PC

STREET: 805 Third Avenue
CITY: New York

STATE: New York
COUNTRY: USA

ZIP: 10022

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/09/500,811
FILING DATE:

CLASSIFICATION:
PRIOR APPLICATION DATA:

APPLICATION NUMBER: 08/187,319
FILING DATE:

ATTORNEY/AGENT INFORMATION:
NAME: Gogoris, Adda C

REGISTRATION NUMBER: 29,714
REFERENCE/DOCKET NUMBER: 4305/08701

TELECOMMUNICATION INFORMATION:
TELEPHONE: 212-527-7700

TELEFAX: 212-753-6237
TELEX: 236687

INFORMATION FOR SEQ ID NO: 18:
SEQUENCE CHARACTERISTICS:

LENGTH: 1341 amino acids
TYPE: amino acid

TOPOLOGY: linear
MOLECULE TYPE: protein

ORGANISM: Homo sapiens
IMMEDIATE SOURCE:

CLONE: COLLAGEN ALPHA 1 (I)
US-09-500-811-18

Query Match

97.0%; Score 1149; DB 4; Length 1341;

Query Match 97.0%; Score 1149; DB 4; Length 1341;
Best Local Similarity 93.6%; Pred. No. 2e-72;
Matches 206; Conservative 0; Mismatches 2; Indels 12; Gaps 1;

QY 1 GPP-----GEGPTGLPGPPGGRGPGSGRPGADGVAGPKGPAGRGSPGPA 48
DB 334 GPPGPAEGEGKRGARGEPTGLPGPPGGRGPGSGRPGADGVAGPKGPAGRGSPGPA 393
QY 49 GPKSGPGEAGRPFGEAGLFGAKGLTGSPGSPGPDGKTGPPGAGQDGRPPGPPGARGQA 108
DB 394 GPKSGPGEAGRPFGEAGLFGAKGLTGSPGSPGPDGKTGPPGAGQDGRPPGPPGARGQA 453
QY 109 GVMGPPGPKGAAGRPFGEAGLFGAKGLTGSPGSPGPDGKTGPPGAGQDGRPPGPPGARGQA 168
DB 454 GVMGPPGPKGAAGRPFGEAGLFGAKGLTGSPGSPGPDGKTGPPGAGQDGRPPGPPGARGQA 513
QY 169 GSPFGQGLPGPAGPPGEAGKPGEGQVPGDLAGPSPGARG 208
DB 514 GSPFGQGLPGPAGPPGEAGKPGEGQVPGDLAGPSPGARG 553

RESULT 3

US-09-570-573-18

; Sequence 18, Application US/09570573
; Patent No. 6342361

GENERAL INFORMATION:
; APPLICANT: Qvist, Per

APPLICANT: Bonde, Martin
; TITLE OF INVENTION: A Method for Assaying Collagen Fragments

; TITLE OF INVENTION: in Body Fluids, A Test Kit and Means for Carrying Out the
; TITLE OF INVENTION: Method and Use of the Method to Diagnose the Presence of
; TITLE OF INVENTION: Disorders Associated with the Metabolism of
; NUMBER OF SEQUENCES: 21

CORRESPONDENCE ADDRESS:
ADDRESSEE: Darby & Darby PC

STREET: 805 Third Avenue
CITY: New York

STATE: New York
COUNTRY: USA

ZIP: 10022

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/09/570,573
FILING DATE:

CLASSIFICATION:
PRIOR APPLICATION DATA:

APPLICATION NUMBER: 08/187,319
FILING DATE:

ATTORNEY/AGENT INFORMATION:
NAME: Gogoris, Adda C

REGISTRATION NUMBER: 29,714
REFERENCE/DOCKET NUMBER: 4305/08701

TELECOMMUNICATION INFORMATION:
TELEPHONE: 212-527-7700

TELEFAX: 212-753-6237
TELEX: 236687

INFORMATION FOR SEQ ID NO: 18:
SEQUENCE CHARACTERISTICS:

LENGTH: 1341 amino acids
TYPE: amino acid

TOPOLOGY: linear
MOLECULE TYPE: protein

ORGANISM: Homo sapiens
IMMEDIATE SOURCE:

CLONE: COLLAGEN ALPHA 1 (I)
US-09-570-573-18

Best Local Similarity 93.6%; Pred. No. 2e-72; Mismatches 2; Indels 12; Gaps 1;
Matches 206; Conservative 0;

QY 1 GPP-----GEPGTGLPDPGGRGSGRFFPCADGVAGPKGPGAGSGPGPA 48
DB 334 GPPGAGEGKRGAGGPGTGLPDPGGRGSGRFFPCADGVAGPKGPGAGSGPGPA 393

QY 49 GPKGSPGAEAGRPAGLPGAKGLTSGSPGDPKGTTPPGACQDGRPPGPPGARGQA 108
DB 394 GPKGSPGAEAGRPAGLPGAKGLTSGSPGDPKGTTPPGACQDGRPPGPPGARGQA 453

QY 109 GVMGFPKGAAGPAGKAGRGVPPGAVGPGAGKDGAGAGAGQPPGAGPAGERGEGGPA 168
DB 454 GVMGFPKGAAGPAGKAGRGVPPGAVGPGAGKDGAGAGAGQPPGAGPAGERGEGGPA 513

QY 169 GSPGFQGLPAGPAGKAGRGVPPGAVGPGAGKDGAGAGAGQPPGAGPAGERGEGGPA 208
DB 514 GSPGFQGLPAGPAGKAGRGVPPGAVGPGAGKDGAGAGAGQPPGAGPAGERGEGGPA 553

RESULT 4
US-09-548-608-18
; Sequence 18, Application US/09548608
; Patent No. 6355442
; GENERAL INFORMATION:
; APPLICANT: Qvist, Per
; APPLICANT: Bonde, Martin
; TITLE OF INVENTION: A Method for Assaying Collagen Fragments
; TITLE OF INVENTION: In Body Fluids, A Test Kit and Means for Carrying Out the
; TITLE OF INVENTION: Method and Use of the Method to Diagnose the Presence of
; TITLE OF INVENTION: Disorders Associated with the Metabolism of
; NUMBER OF SEQUENCES: 21
; CORRESPONDENCE ADDRESSES:
; ADDRESSEE: Darby & Darby PC
; STREET: 805 Third Avenue
; CITY: New York
; STATE: New York
; COUNTRY: USA
; ZIP: 10022
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent in Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/548,608
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/187,319
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Gogoris, Adda C
; REGISTRATION NUMBER: 29,714
; REFERENCE/DOCKET NUMBER: 4305/08701
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 212-527-7700
; TELEFAX: 212-753-6237
; TELEX: 236687
; INFORMATION FOR SEQ ID NO: 18:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1341 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; ORIGINAL SOURCE:
; ORGANISM: Homo sapiens
; IMMEDIATE SOURCE:
; CLONE: COLLAGEN ALPHA 1 (I)

Query Match 97.0%; Score 1149; DB 4; Length 1341;
Best Local Similarity 93.6%; Pred. No. 2e-72;

Matches 206; Conservative 0; Mismatches 2; Indels 12; Gaps 1;
QY 1 GPP-----GEPGTGLPDPGGRGSGRFFPCADGVAGPKGPGAGSGPGPA 48
DB 334 GPPGAGEGKRGAGGPGTGLPDPGGRGSGRFFPCADGVAGPKGPGAGSGPGPA 393

QY 49 GPKGSPGAEAGRPAGLPGAKGLTSGSPGDPKGTTPPGACQDGRPPGPPGARGQA 108
DB 394 GPKGSPGAEAGRPAGLPGAKGLTSGSPGDPKGTTPPGACQDGRPPGPPGARGQA 453

QY 109 GVMGFPKGAAGPAGKAGRGVPPGAVGPGAGKDGAGAGAGQPPGAGPAGERGEGGPA 168
DB 454 GVMGFPKGAAGPAGKAGRGVPPGAVGPGAGKDGAGAGAGQPPGAGPAGERGEGGPA 513

QY 169 GSPGFQGLPAGPAGKAGRGVPPGAVGPGAGKDGAGAGAGQPPGAGPAGERGEGGPA 208
DB 514 GSPGFQGLPAGPAGKAGRGVPPGAVGPGAGKDGAGAGAGQPPGAGPAGERGEGGPA 553

RESULT 5
US-09-585-887-9
; Sequence 9, Application US/09585887
; Patent No. 6413742
; GENERAL INFORMATION:
; APPLICANT: Olsen, David R
; APPLICANT: Chang, Robert
; APPLICANT: McMullin, Hugh
; APPLICANT: Hitzeman, Ronald A.
; APPLICANT: Chisholm, George
; TITLE OF INVENTION: NOVEL METHODS FOR THE PRODUCTION OF GELATIN AND
; TITLE OF INVENTION: FULL-LENGTH TRIPLE HELICAL COLLAGEN IN RECOMBINANT
; TITLE OF INVENTION: CELLS
; FILE REFERENCE: 225002030400
; CURRENT APPLICATION NUMBER: US/09/585,887
; CURRENT FILING DATE: 2000-05-31
; PRIOR APPLICATION NUMBER: 09/289,578
; PRIOR FILING DATE: 1999-04-09
; PRIOR APPLICATION NUMBER: 60/084,828
; PRIOR FILING DATE: 1998-05-08
; NUMBER OF SEQ ID NOS: 11
; SOFTWARE: Patent in Ver. 2.0
; SEQ ID NO: 9
; LENGTH: 1461
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-585-887-9

Query Match 97.0%; Score 1149; DB 4; Length 1461;
Best Local Similarity 93.6%; Pred. No. 2.2e-72;
Matches 206; Conservative 0; Mismatches 2; Indels 12; Gaps 1;
QY 1 GPP-----GEPGTGLPDPGGRGSGRFFPCADGVAGPKGPGAGSGPGPA 48
DB 455 GPPGAGEGKRGAGGPGTGLPDPGGRGSGRFFPCADGVAGPKGPGAGSGPGPA 514

QY 49 GPKGSPGAEAGRPAGLPGAKGLTSGSPGDPKGTTPPGACQDGRPPGPPGARGQA 108
DB 515 GPKGSPGAEAGRPAGLPGAKGLTSGSPGDPKGTTPPGACQDGRPPGPPGARGQA 574

QY 109 GVMGFPKGAAGPAGKAGRGVPPGAVGPGAGKDGAGAGAGQPPGAGPAGERGEGGPA 168
DB 575 GVMGFPKGAAGPAGKAGRGVPPGAVGPGAGKDGAGAGAGQPPGAGPAGERGEGGPA 634

QY 169 GSPGFQGLPAGPAGKAGRGVPPGAVGPGAGKDGAGAGAGQPPGAGPAGERGEGGPA 208
DB 635 GSPGFQGLPAGPAGKAGRGVPPGAVGPGAGKDGAGAGAGQPPGAGPAGERGEGGPA 674

RESULT 6
US-09-289-578-9
; Sequence 9, Application US/09289578
; Patent No. 6428978
; GENERAL INFORMATION:

; APPLICANT: Olsen, David R
 ; APPLICANT: Chang, Robert
 ; APPLICANT: McMullin, Hugh
 ; APPLICANT: Hitzeman, Ronald A.
 ; APPLICANT: Chisholm, George
 ; TITLE OF INVENTION: NOVEL METHODS FOR THE PRODUCTION OF GELATIN AND
 ; TITLE OF INVENTION: FULL-LENGTH TRIPLE HELICAL COLLAGEN IN RECOMBINANT
 ; TITLE OF INVENTION: CELLS
 ; FILE REFERENCE: 225002030400
 ; CURRENT APPLICATION NUMBER: US/09/289,578
 ; CURRENT FILING DATE: 1999-04-10
 ; PRIOR APPLICATION NUMBER: 60/084,828
 ; PRIOR FILING DATE: 1998-05-08
 ; NUMBER OF SEQ ID NOS: 11
 ; SOFTWARE: PatentIn Ver. 2.0
 ; SEQ ID NO 9
 ; LENGTH: 1461
 ; TYPE: PRT
 ; ORGANISM: Homo sapiens
 US-09-289-578-9

Query Match 97.0%; Score 1149; DB 4; Length 1461;
 Best Local Similarity 93.6%; Pred. No. 2.2e-72;
 Matches 206; Conservative 0; Mismatches 2; Indels 12; Gaps 1;

QY 1 GPP-----GEGPTGLPGPPGGRGGSGRFFGADGVAGPKPAGERGSPGA 48
 DB 455 GPPGAGBEKGKARGEGPTGLPGPPGGRGGSGRFFGADGVAGPKPAGERGSPGA 514
 QY 49 GPKGSPGAGRPGEAGLPGAAGLGTGSPGSPGDPCKTGGPPAGQDGRPGPPPGARGQA 108
 DB 515 GPKGSPGAGRPGEAGLPGAAGLGTGSPGSPGDPCKTGGPPAGQDGRPGPPPGARGQA 574
 QY 109 GWMGPPGPKGAAGEPKKAGERGVPPGCAVGPAGKDGEGAGAGPPGAPGAGERGEGQA 168
 DB 575 GWMGPPGPKGAAGEPKKAGERGVPPGCAVGPAGKDGEGAGAGPPGAPGAGERGEGQA 634
 QY 169 GSGFGQGLPGAPPPGEAGKPGEGQVPGDLGAPGSPAG 208
 DB 635 GSGFGQGLPGAPPPGEAGKPGEGQVPGDLGAPGSPAG 674

RESULT 7
 US-09-331-347C-21
 ; Sequence 21, Application US/09331347C
 ; Patent No. 6617431
 ; GENERAL INFORMATION:
 ; APPLICANT: Meristem Therapeutics, S.A.
 ; APPLICANT: Meristem Therapeutics, S.A.
 ; TITLE OF INVENTION: Recombinant Collagens and Derived Proteins Produced by P
 ; TITLE OF INVENTION: obtaining Such and Their Uses
 ; FILE REFERENCE: 1149-3
 ; CURRENT APPLICATION NUMBER: US/09/331.347C
 ; CURRENT FILING DATE: 1999-08-17
 ; NUMBER OF SEQ ID NOS: 22
 ; SOFTWARE: PatentIn version 3.1
 ; SEQ ID NO 21
 ; LENGTH: 1464
 ; TYPE: PRT
 ; ORGANISM: Homo sapiens
 US-09-331-347C-21

Query Match 97.0%; Score 1149; DB 4; Length 1464;
 Best Local Similarity 93.6%; Pred. No. 2.2e-72;
 Matches 206; Conservative 0; Mismatches 2; Indels 12; Gaps 1;

QY 1 GPP-----GEGPTGLPGPPGGRGGSGRFFGADGVAGPKPAGERGSPGA 48
 DB 458 GPPGAGBEKGKARGEGPTGLPGPPGGRGGSGRFFGADGVAGPKPAGERGSPGA 517
 QY 49 GPKGSPGAGRPGEAGLPGAAGLGTGSPGSPGDPCKTGGPPAGQDGRPGPPPGARGQA 108
 DB 518 GPKGSPGAGRPGEAGLPGAAGLGTGSPGSPGDPCKTGGPPAGQDGRPGPPPGARGQA 577

```

Qy 109 GVMGPPGPKGAAGBPGKAGERGVPVPGAVGPPAGKDGGAAGQPPGDPAGPAGERGEQGPA 163
Db 578 GVMGPPGPKGAAGBPGKAGERGVPVPGAVGPPAGKDGGAAGQPPGDPAGPAGERGEQGPA 637
Qy 169 GSPGFQGLPGAGPPGAGKRGPEQGVPCDLGAPGSPGAG 208
Db 638 GSPGFQGLPGAGPPGAGKRGPEQGVPCDLGAPGSPGAG 677

RESULT 8
US-08-931-820-1
; Sequence 1, Application US/08931820
; Patent No. 6010863
; GENERAL INFORMATION:
; APPLICANT:
; TITLE OF INVENTION: Assay for collagen degradation
; NUMBER OF SEQUENCES: 4
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25 (EPO)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/931,820
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: EP 96202596.1
; FILING DATE:
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1057 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; HYPOTHEICAL: NO
; ANTI-SENSE: NO
; ORIGINAL SOURCE:
; ORGANISM: Homo sapiens
; TISSUE TYPE: Collagen type I
US-08-931-820-1

Query Match 94.6%; Score 1120; DB 3; Length 1057;
Best Local Similarity 96.2%; Pred. No. 1.7e-70;
Matches 200; Conservative 0; Mismatches 8; Indels 0; Gaps 0;

Qy 1 GPPGEPGPTGLPGPPGRRGGPGRGFFGADGVAGPKGAPAGERGSPGAPGKSGPGEAGRP 60
Db 309 GPNCEAGSAGPPGPPGLRGFGSRGFFGADGVAGPKGAPAGERGSPGAPGKSGPGEAGRP 368
Qy 61 GEAGLPAGKAGLTGSPGSPGDDGKTGPPGPAQDGRPPGPPGASGQAGVMGFFGPKGAA 120
Db 369 GEAGLPAGKAGLTGSPGSPGDDGKTGPPGPAQDGRPPGPPGASGQAGVMGFFGPKGAA 428
Qy 121 GEPGKAGERGVPVPGAVGPPAGKDGGAAGQPPGPPAGPAGERGEQGPAGSPGFCQLPGPA 180
Db 429 GEPGKAGERGVPVPGAVGPPAGKDGGAAGQPPGPPAGPAGERGEQGPAGSPGFCQLPGPA 488
Qy 181 GPPGAGKRGPEQGVPCDLGAPGSPGAG 208
Db 489 GPPGAGKRGPEQGVPCDLGAPGSPGAG 516

RESULT 9
US-09-219-849-48
; Sequence 48, Application US/09219849
; Patent No. 615081
; GENERAL INFORMATION:
; APPLICANT: VAN HEERDE, GEORGE V.
; APPLICANT: VAN RIJN, ALEXIS C.
; APPLICANT: BOWSTRA, JAN B.

```

```

; APPLICANT: DE WOLF, FREDERIK A.
; APPLICANT: MOOBROEK, ANDREAS
; APPLICANT: WERTEN, MARC W.T.
; APPLICANT: WIND, RICHEL D.
; APPLICANT: VAN DEN BOSCH, TANJA J.
; TITLE OF INVENTION: SILVER HALIDE EMULSIONS WITH RECOMBINANT COLLAGEN
; TITLE OF INVENTION: SUITABLE FOR PHOTOGRAPHIC APPLICATION AND ALSO THE
; TITLE OF INVENTION: PREPARATION THEREOF
; FILE REFERENCE: 2728-2
; CURRENT APPLICATION NUMBER: US/09/219,849
; CURRENT FILING DATE: 1998-12-23
; NUMBER OF SEQ ID NOS: 50
; SOFTWARE: Patent In Ver. 2.1
; SEQ ID NO 48
; LENGTH: 595
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: Synthetic
; OTHER INFORMATION: amino acid sequence
US-09-219-849-48

Query Match          94.3%; Score 1116; DB 3; Length 595;
Best Local Similarity 90.5%; Pred. No. 1.9e-70;
Matches 199; Conservative 4; Mismatches 5; Indels 12; Gaps 1;

QY 1 GPP-----GEPGPTGLPPEERGGPSRGGPGADGVAGPKGPGAGSGSPGA 48
DB 279 GPPPAGEEGRKARGEPGSLGPPGERGGPSRGGPGADGVAGPKGPGAGSGPA 338
QY 49 GPKSGPEAGRPGEAGLPGAKGLTSGSPGPDGKTGPPGAGQDGRPPGPPGARGQA 108
DB 339 GPKSGPEAGRPGEAGLPGAKGLTSGSPGPDGKTGPPGAGQDGRPPGPPGARGQA 398
QY 109 GVMGFPFGKGAAGEPGKAGRGVPPGAVGPGAGKDGAGAGQDGRPPGAGRGEGQGA 168
DB 399 GVMGFPFGKGTAGBPGKAGRGVPPGAVGPGAGKDGAGAGQDGRPPGAGRGEGQGA 458
QY 169 GSPGFQGLPGAGPPGEAGKPGEGQGVPGDLGAPGSPGAG 208
DB 459 GSPGFQGLPGAGPPGEAGKPGEGQGVPGDLGAPGSPGARG 498

RESULT 10
US-09-219-849-50
; Sequence 50, Application US/09219849
; Patent No. 6150081
; GENERAL INFORMATION:
; APPLICANT: VAN HEERDE, GEORGE V.
; APPLICANT: VAN RIJN, ALEXIS C.
; APPLICANT: BOWMSTRA, JAN B.
; APPLICANT: DE WOLF, FREDERIK A.
; APPLICANT: MOOBROEK, ANDREAS
; APPLICANT: WERTEN, MARC W.T.
; APPLICANT: WIND, RICHEL D.
; APPLICANT: VAN DEN BOSCH, TANJA J.
; TITLE OF INVENTION: SILVER HALIDE EMULSIONS WITH RECOMBINANT COLLAGEN
; TITLE OF INVENTION: SUITABLE FOR PHOTOGRAPHIC APPLICATION AND ALSO THE
; TITLE OF INVENTION: PREPARATION THEREOF
; FILE REFERENCE: 2728-2
; CURRENT APPLICATION NUMBER: US/09/219,849
; CURRENT FILING DATE: 1998-12-23
; NUMBER OF SEQ ID NOS: 50
; SOFTWARE: Patent In Ver. 2.1
; SEQ ID NO 49
; LENGTH: 822
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: Synthetic
; OTHER INFORMATION: amino acid sequence
US-09-219-849-49

Query Match          94.3%; Score 1116; DB 3; Length 822;
Best Local Similarity 90.5%; Pred. No. 2.6e-70;
Matches 199; Conservative 4; Mismatches 5; Indels 12; Gaps 1;

QY 1 GPP-----GEPGPTGLPPEERGGPSRGGPGADGVAGPKGPGAGSGSPGA 48
DB 279 GPPPAGEEGRKARGEPGSLGPPGERGGPSRGGPGADGVAGPKGPGAGSGPA 338
QY 49 GPKSGPEAGRPGEAGLPGAKGLTSGSPGPDGKTGPPGAGQDGRPPGPPGARGQA 108
DB 339 GPKSGPEAGRPGEAGLPGAKGLTSGSPGPDGKTGPPGAGQDGRPPGPPGARGQA 398
QY 109 GVMGFPFGKGAAGEPGKAGRGVPPGAVGPGAGKDGAGAGQDGRPPGAGRGEGQGA 168
DB 399 GVMGFPFGKGTAGBPGKAGRGVPPGAVGPGAGKDGAGAGQDGRPPGAGRGEGQGA 458
QY 169 GSPGFQGLPGAGPPGEAGKPGEGQGVPGDLGAPGSPGAG 208
DB 459 GSPGFQGLPGAGPPGEAGKPGEGQGVPGDLGAPGSPGARG 498

RESULT 11
US-09-219-849-49
; Sequence 49, Application US/09219849
; Patent No. 6150081
; GENERAL INFORMATION:
; APPLICANT: VAN HEERDE, GEORGE V.
; APPLICANT: VAN RIJN, ALEXIS C.
; APPLICANT: BOWMSTRA, JAN B.
; APPLICANT: DE WOLF, FREDERIK A.
; APPLICANT: MOOBROEK, ANDREAS
; APPLICANT: WERTEN, MARC W.T.
; APPLICANT: WIND, RICHEL D.
; APPLICANT: VAN DEN BOSCH, TANJA J.
; TITLE OF INVENTION: SILVER HALIDE EMULSIONS WITH RECOMBINANT COLLAGEN
; TITLE OF INVENTION: SUITABLE FOR PHOTOGRAPHIC APPLICATION AND ALSO THE
; TITLE OF INVENTION: PREPARATION THEREOF
; FILE REFERENCE: 2728-2
; CURRENT APPLICATION NUMBER: US/09/219,849
; CURRENT FILING DATE: 1998-12-23
; NUMBER OF SEQ ID NOS: 50
; SOFTWARE: Patent In Ver. 2.1
; SEQ ID NO 49
; LENGTH: 822
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: Synthetic
; OTHER INFORMATION: amino acid sequence
US-09-219-849-49

Query Match          94.3%; Score 1116; DB 3; Length 822;
Best Local Similarity 90.5%; Pred. No. 2.6e-70;
Matches 199; Conservative 4; Mismatches 5; Indels 12; Gaps 1;

QY 1 GPP-----GEPGPTGLPPEERGGPSRGGPGADGVAGPKGPGAGSGSPGA 48
DB 279 GPPPAGEEGRKARGEPGSLGPPGERGGPSRGGPGADGVAGPKGPGAGSGPA 338
QY 49 GPKSGPEAGRPGEAGLPGAKGLTSGSPGPDGKTGPPGAGQDGRPPGPPGARGQA 108
DB 339 GPKSGPEAGRPGEAGLPGAKGLTSGSPGPDGKTGPPGAGQDGRPPGPPGARGQA 398
QY 109 GVMGFPFGKGAAGEPGKAGRGVPPGAVGPGAGKDGAGAGQDGRPPGAGRGEGQGA 168
DB 399 GVMGFPFGKGTAGBPGKAGRGVPPGAVGPGAGKDGAGAGQDGRPPGAGRGEGQGA 458
QY 169 GSPGFQGLPGAGPPGEAGKPGEGQGVPGDLGAPGSPGAG 208
DB 459 GSPGFQGLPGAGPPGEAGKPGEGQGVPGDLGAPGSPGARG 498

RESULT 12
```

```

; APPLICANT: DE WOLF, FREDERIK A.
; APPLICANT: MOOBROEK, ANDREAS
; APPLICANT: WERTEN, MARC W.T.
; APPLICANT: WIND, RICHEL D.
; APPLICANT: VAN DEN BOSCH, TANJA J.
; TITLE OF INVENTION: SILVER HALIDE EMULSIONS WITH RECOMBINANT COLLAGEN
; TITLE OF INVENTION: SUITABLE FOR PHOTOGRAPHIC APPLICATION AND ALSO THE
; TITLE OF INVENTION: PREPARATION THEREOF
; FILE REFERENCE: 2728-2
; CURRENT APPLICATION NUMBER: US/09/219,849
; CURRENT FILING DATE: 1998-12-23
; NUMBER OF SEQ ID NOS: 50
; SOFTWARE: Patent In Ver. 2.1
; SEQ ID NO 48
; LENGTH: 595
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: Synthetic
; OTHER INFORMATION: amino acid sequence
US-09-219-849-48

Query Match          94.3%; Score 1116; DB 3; Length 595;
Best Local Similarity 90.5%; Pred. No. 1.9e-70;
Matches 199; Conservative 4; Mismatches 5; Indels 12; Gaps 1;

QY 1 GPP-----GEPGPTGLPPEERGGPSRGGPGADGVAGPKGPGAGSGSPGA 48
DB 279 GPPPAGEEGRKARGEPGSLGPPGERGGPSRGGPGADGVAGPKGPGAGSGPA 338
QY 49 GPKSGPEAGRPGEAGLPGAKGLTSGSPGPDGKTGPPGAGQDGRPPGPPGARGQA 108
DB 339 GPKSGPEAGRPGEAGLPGAKGLTSGSPGPDGKTGPPGAGQDGRPPGPPGARGQA 398
QY 109 GVMGFPFGKGAAGEPGKAGRGVPPGAVGPGAGKDGAGAGQDGRPPGAGRGEGQGA 168
DB 399 GVMGFPFGKGTAGBPGKAGRGVPPGAVGPGAGKDGAGAGQDGRPPGAGRGEGQGA 458
QY 169 GSPGFQGLPGAGPPGEAGKPGEGQGVPGDLGAPGSPGAG 208
DB 459 GSPGFQGLPGAGPPGEAGKPGEGQGVPGDLGAPGSPGARG 498

RESULT 10
US-09-219-849-50
; Sequence 50, Application US/09219849
; Patent No. 6150081
; GENERAL INFORMATION:
; APPLICANT: VAN HEERDE, GEORGE V.
; APPLICANT: VAN RIJN, ALEXIS C.
; APPLICANT: BOWMSTRA, JAN B.
; APPLICANT: DE WOLF, FREDERIK A.
; APPLICANT: MOOBROEK, ANDREAS
; APPLICANT: WERTEN, MARC W.T.
; APPLICANT: WIND, RICHEL D.
; APPLICANT: VAN DEN BOSCH, TANJA J.
; TITLE OF INVENTION: SILVER HALIDE EMULSIONS WITH RECOMBINANT COLLAGEN
; TITLE OF INVENTION: SUITABLE FOR PHOTOGRAPHIC APPLICATION AND ALSO THE
; TITLE OF INVENTION: PREPARATION THEREOF
; FILE REFERENCE: 2728-2
; CURRENT APPLICATION NUMBER: US/09/219,849
; CURRENT FILING DATE: 1998-12-23
; NUMBER OF SEQ ID NOS: 50
; SOFTWARE: Patent In Ver. 2.1
; SEQ ID NO 50
; LENGTH: 595
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: Synthetic
; OTHER INFORMATION: amino acid sequence
US-09-219-849-50
```



```
SEQUENCE CHARACTERISTICS:
LENGTH: 1060 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
HYPOTHETICAL: NO
ORIGINAL SOURCE:
ORGANISM: Homo sapiens
TISSUE TYPE: Collagen type II
US-08-931-820-3

Query Match      72.8%; Score 862; DB 3; Length 1060;
Best Local Similarity 74.0%; Pred. No. 1.3e-52;
Matches 154; Conservative 11; Mismatches 43; Indels 0; Gaps 0;

QY 1 GPPGEPPTGLPDPGPPGSGRGGFCADGVAGPKGPAAGERSGPPGAGPKGSPGFEAGRP 60
DB 311 GARGEFGVGPFGPPGGERGAPGNRFFGQDGLAGPKGAPGERGSGLAGPKGANGDFGRP 370
QY 61 GEAGLPAGKLTGSGSPGPGDKTGPAGQDGRPPGPPGARGQAGVMGFPKGA 120
DB 371 GEPGLPARGLTGRPDAGPQGVPSGAPGEDGRPPGQARGQGVGMGFPKGAN 430
QY 121 GEPKAGRGVPPGPAVGPAGKDGAGAGQPPGPAAGRGPGAGSPGFGQLPAPA 180
DB 431 GEPKAGRGVPPGPAVGPAGKDGAGAGQPPGPAAGRGPGAGRGQAGAPGSGFQLPAPP 490
QY 181 GPPGAGKPGGQGVPGDLGAPSPGAP 208
DB 491 GPPGEGGKPGDQGVGEGAGAPGLVPRG 518

RESULT 15
US-08-963-825-20
Sequence 20, Application US/08963825
Patent No. 6110689
GENERAL INFORMATION:
APPLICANT: Qvist, Per
APPLICANT: Bonde, Martin
TITLE OF INVENTION: A Method for Assaying Collagen Fragments
TITLE OF INVENTION: In Body Fluids, A Test Kit and Means for Carrying Out the
TITLE OF INVENTION: Method and Use of the Method to Diagnose the Presence of
TITLE OF INVENTION: Disorders Associated with the Metabolism of
NUMBER OF SEQUENCES: 21
CORRESPONDENCE ADDRESS:
ADDRESSEE: Darby & Darby PC
STREET: 805 Third Avenue
CITY: New York
STATE: New York
COUNTRY: USA
ZIP: 10022
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/963,825
FILING DATE:
CLASSIFICATION: 436
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/08/187,319
FILING DATE: 21-JAN-1994
ATTORNEY/AGENT INFORMATION:
NAME: Gogoris, Adda C
REGISTRATION NUMBER: 29,714
REFERENCE/DOCKET NUMBER: 4305/08701
TELECOMMUNICATION INFORMATION:
TELEPHONE: 212-527-7700
TELEFAX: 212-753-6237
TELEX: 236687
INFORMATION FOR SEQ ID NO: 20:
```

```
SEQUENCE CHARACTERISTICS:
LENGTH: 1418 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
ORIGINAL SOURCE:
ORGANISM: Homo sapiens
IMMEDIATE SOURCE:
CLONE: COLLAGEN -ALPHA 1 (II)
US-08-963-825-20

Query Match      72.8%; Score 862; DB 3; Length 1418;
Best Local Similarity 74.0%; Pred. No. 1.7e-52;
Matches 154; Conservative 11; Mismatches 43; Indels 0; Gaps 0;

QY 1 GPPGEPPTGLPDPGPPGSGRGGFCADGVAGPKGPAAGERSGPPGAGPKGSPGFEAGRP 60
DB 423 GARGEFGVGPFGPPGGERGAPGNRFFGQDGLAGPKGAPGERGSGLAGPKGANGDFGRP 482
QY 61 GEAGLPAGKLTGSGSPGPGDKTGPAGQDGRPPGPPGARGQAGVMGFPKGA 120
DB 483 GEPGLPARGLTGRPDAGPQGVPSGAPGEDGRPPGQARGQGVGMGFPKGAN 542
QY 121 GEPKAGRGVPPGPAVGPAGKDGAGAGQPPGPAAGRGPGAGRGQAGAPGSGFQLPAPA 180
DB 543 GEPKAGRGVPPGPAVGPAGKDGAGAGQPPGPAAGRGPGAGRGQAGAPGSGFQLPAPP 602
QY 181 GPPGAGKPGGQGVPGDLGAPSPGAP 208
DB 603 GPPGEGGKPGDQGVGEGAGAPGLVPRG 630

Search completed: September 24, 2004, 11:09:49
Job time : 10.6042 secs
```

This Page Blank (uspto)

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: September 24, 2004, 11:06:55 ; Search time 28.5883 Seconds
(without alignments)
2065.614 Million cell updates/sec

Title: US-10-658-989a-1

Perfect score: 1184

Sequence: 1 GPPGEPGPTGLPQPPGRRG.....GEGVFDLCPGSPGAG 209

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1586107 seqs, 282547505 residues

Total number of hits satisfying chosen parameters: 1586107

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : A_Geneseq_29Jan04.*

- 1: Geneseqp1980s.*
- 2: Geneseqp1980s.*
- 3: Geneseqp2000s.*
- 4: Geneseqp2001s.*
- 5: Geneseqp2002s.*
- 6: Geneseqp2003as.*
- 7: Geneseqp2003bs.*
- 8: Geneseqp2004s.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1149	97.0	501	4 AAB68057	Aab68057 Amino aci
2	1149	97.0	501	4 AAB02703	Aae02703 Human alp
3	1149	97.0	501	7 ADB84290	Adb84290 Recombina
4	1149	97.0	1057	3 AAY84541	Aay84541 Amino aci
5	1149	97.0	1057	3 AAY84544	Aay84544 A human c
6	1149	97.0	1058	3 AAY84403	Aay84403 Amino aci
7	1149	97.0	1107	2 AAR89472	Aar89472 Collagen/
8	1149	97.0	1107	3 AAY84540	Aay84540 Amino aci
9	1149	97.0	1161	7 ADE87050	Ade87050 Human pan
10	1149	97.0	1169	2 AAR89469	Aar89469 Collagen/
11	1149	97.0	1169	3 AAY84537	Aay84537 Amino aci
12	1149	97.0	1171	2 AAR89470	Aar89470 Collagen/
13	1149	97.0	1171	3 AAY84538	Aay84538 A chimeri
14	1149	97.0	1211	7 ADE87057	Ade87057 Human pan
15	1149	97.0	1226	7 ADE87062	Ade87062 Human pan
16	1149	97.0	1341	2 AAR71701	Aar71701 Collagen
17	1149	97.0	1341	3 AAY96122	Aay96122 Collagen
18	1149	97.0	1341	5 AAE16475	Aae16475 Human col
19	1149	97.0	1341	5 ABB080733	Abb080733 Collagen
20	1149	97.0	1341	5 ABB09625	Abb09625 Amino aci
21	1149	97.0	1388	2 AAR89471	Aar89471 Collagen/
22	1149	97.0	1411	3 AAY56800	Aay56800 Human pre
23	1149	97.0	1461	5 ABG93947	Abg93947 Human pol
24	1149	97.0	1464	2 AAW68485	Aaw68485 Human rec
25	1149	97.0	1464	4 AAB82454	Aab82454 Human pro

26	1149	97.0	1464	4 AAB14136	Aab14136 Human nov
27	1149	97.0	1464	5 ABB90764	Abb90764 Human tum
28	1149	97.0	1464	5 ABB90764	Abb90764 Human pan
29	1149	97.0	1464	6 ABUS4471	Abus4471 Human tum
30	1149	97.0	1464	6 ABR47417	Abx47417 Breast ca
31	1149	97.0	1464	6 ABR92064	Abx92064 Human cer
32	1149	97.0	1464	7 ADD14142	Add14142 Human src
33	1149	97.0	1464	7 ADD45059	Add45059 Human pro
34	1149	97.0	1464	7 ADD45055	Add45055 Human pro
35	1149	97.0	1464	7 ADD45051	Add45051 Human pro
36	1149	97.0	1464	7 ADE87048	Ade87048 Human pan
37	1149	97.0	1536	7 ADE87051	Ade87051 Human pan
38	1144	96.6	1449	4 AAE02535	Aae02535 Porcine a
39	1141	95.4	1388	3 AAY84539	Aay84539 Amino aci
40	1141	95.4	1463	4 AAE02532	Aae02532 Bovine al
41	1128	95.3	1518	4 ABG22679	Abg22679 Novel hum
42	1122	94.8	1453	7 ADD45053	Add45053 Rat Prote
43	1122	94.8	1453	7 ADD45057	Add45057 Rat Prote
44	1122	94.8	1453	7 ADD48341	Add48341 Rat Prote
45	1122	94.8	1453	7 ADD45049	Add45049 Rat Prote

ALIGNMENTS

RESULT 1

AAB68057

ID AAB68057 standard; protein; 501 AA.

XX AAB68057;

AC AAB68057;

DT 09-JUL-2001 (first entry)

XX Amino acid sequence of a recombinant human gelatin.

DE Human; gelatin; vaccine; anaphylactic reaction.

KW Homo sapiens.

XX Key

XX Location/Qualifiers

FT Misc-difference 85

FT /note= "this residue is given as unknown as it is

FT illegible in the specification"

XX WO200134801-A2.

XX 17-MAY-2001.

XX 10-NOV-2000; 2000WO-US030843.

XX 12-NOV-1999; 99US-0165114P.

XX 15-MAY-2000; 2000US-0204437P.

XX (FIBR-) FIBROGEN INC.

XX Chang RC, Kivirikko KI, Neff TB, Olsen DR, Polarek JW;

XX WPI; 2001-308784/32.

XX Vaccine formulations (i) comprising recombinant human gelatin, useful for
vaccinating against e.g. mumps, measles, rubella, tetanus, rabies and
cholera, the gelatin is non-immunogenic and confers stability at ambient
temperatures.

XX Claim 11; Page 114-116; 130pp; English.

XX The present sequence represents a human recombinant gelatin polypeptide.
The recombinant gelatin polypeptide is used to produce vaccine
formulations of the invention. The recombinant human gelatin is non-
immunogenic (therefore reducing anaphylactic reactions) and confers
stability at ambient temperatures. The vaccine formulation comprises a
vaccine formulated for the prevention of a disease selected from vaccinia
virus (small pox), polio virus (Salk and Sabin), mumps, measles, rubella,

PR 15-MAY-2000; 2000US-0204437P.
 PR 10-NOV-2000; 2000US-00710249.
 XX (CHAN//) CHANG R C.
 PA (KIVI//) KIVIRIKKO K I.
 PA (NEFF//) NEFF T B.
 PA (OLSE//) OLSEN D R.
 PA (POLA//) POLAREK J W.
 XX Chang RC, Kivirikko KI, Neff TB, Olsen DR, Polarek JW;
 XX WPI; 2003-540775/51.
 XX New vaccine composition comprising a recombinant gelatin and an antigenic
 PT agent, useful for preventing e.g. polio virus, mumps, measles, rubella,
 PT diphtheria, tetanus, chicken pox/shingles, pertussis, cholera, rotavirus
 PT or dengue.
 XX Claim 16; Page 36-38; 63pp; English.
 XX The invention describes a vaccine composition comprising a recombinant
 CC gelatin, and an antigenic agent. The vaccine can be delivered by
 CC injection, through nasal, oral, transdermal or mucosal routes, or through
 CC deep lung delivery. Administration may also be oral, rectal,
 CC transcutaneous, intramedullary, intrathecal, intraventricular,
 CC intraperitoneal, intranasal or intracocular injection). The vaccine
 CC composition is formulated for the prevention of vaccinia virus (small
 CC pox), polio virus (Salk and Sabin), mumps, measles, rubella, diphtheria,
 CC tetanus, Varicella-Zoster (chicken pox/shingles), pertussis (whooping
 CC cough), Bacille Calmette-Guerin (BCG, tuberculosis), Haemophilus
 CC influenzae meningitis, rabies, cholera, Japanese encephalitis virus,
 CC Salmonella typhi, Shigella hepatitis A, hepatitis B, adenovirus, yellow
 CC fever, foot and mouth disease, herpes simplex virus, respiratory
 CC syncytial virus, rotavirus, dengue, West Nile virus, Turkey herpes virus
 CC (Marek's disease), influenza, and anthrax. This is the amino acid
 CC sequence of a recombinant gelatin used in the creation of a vaccine
 CC composition of the invention.
 XX Sequence 501 AA;
 XX
 Query Match 97.0%; Score 1149; DB 7; Length 501;
 Best Local Similarity 93.6%; Pred. No. 3.5e-67;
 Matches 206; Conservative 0; Mismatches 2; Indels 12; Gaps 1;
 QY 1 GPP-----GPPGPTGLPGRGSGRFPAGADGVAGPKGAGRGSPGPA 48
 DB 280 GPPGAGEGKRGAGRGFPGLGPPGGRGSGRFPAGADGVAGPKGAGRGSPGPA 339
 QY 49 GPKGSPGAGRPAGLPGAKGLTSGSPGPDGKTGPPAGQDGRPPGPPGARGQA 108
 DB 340 GPKGSPGAGRPAGLPGAKGLTSGSPGPDGKTGPPAGQDGRPPGPPGARGQA 399
 QY 109 GVMGFPKGAGRPAGRGVPGPCAVGPAKDGAGAGQGGPPGAGRGGEQGA 168
 DB 400 GVMGFPKGAGRPAGRGVPGPCAVGPAKDGAGAGQGGPPGAGRGGEQGA 459
 QY 169 GSPGFQGLPGAGPPGAGKFGEGQGVFDLGLGAPGSPGAG 208
 DB 460 GSPGFQGLPGAGPPGAGKFGEGQGVFDLGLGAPGSPGAG 499
 RESULT 4
 AAY84541
 ID AAY84541 standard; protein; 1057 AA.
 XX AAY84541;
 AC AAY84541;
 XX
 DT 25-JUL-2000 (first entry)
 XX
 DE Amino acid sequence of a human collagen 1 (alpha1) protein.
 XX
 KW Extracellular matrix protein; self aggregation; hydroxylated proline;

KW trans-4-hydroxyproline; 3-hydroxyproline; recombinant protein production;
 KW collagen; fibrinogen; fibronectin; post translational hydroxylation.
 OS Homo sapiens.
 XX EP992586-A2.
 XX 12-APR-2000.
 XX 07-OCT-1999; 99EP-00119184.
 XX 09-OCT-1998; 98US-00169768.
 XX (USSU) US SURGICAL CORP.
 XX Gruskin EA, Buechter DD, Zhang G, Connolly K;
 XX WPI; 2000-259138/23.
 XX N-PSDB; AAA12502.
 XX Production of extracellular matrix proteins containing 4-trans-
 PT hydroxyproline results in native self aggregating proteins, useful on
 PT medical implants.
 XX Disclosure; Fig 27A-B; 260pp; English.
 XX The specification describes a method for producing an extracellular
 CC matrix protein or its fragment. The extracellular matrix protein is
 CC capable of self aggregating in a cell which does not ordinarily
 CC hydroxylated prolines. The method comprises optimising a nucleic acid
 CC sequence for expression in the cell by substitution of codons preferred
 CC by that cell for naturally occurring codons not preferred by the cell;
 CC incorporating the nucleic acid sequence into the cell; and contacting the
 CC cell with a hypertonic growth medium containing at least one amino acid,
 CC selected from the group consisting of trans-4-hydroxyproline and 3-
 CC hydroxyproline to allow at least one of the amino acids to be assimilated
 CC into the cell and incorporated into the extracellular matrix protein. The
 CC method may be used to make host cells assimilate and incorporate trans-4-
 CC hydroxyproline into proteins. This is especially useful in the
 CC recombinant production of proteins such as collagen, fibrinogen and
 CC fibronectin whose ability to self aggregate and produce functional
 CC proteins depends on the post translational hydroxylation of proline. The
 CC method is also useful in studying the structure and function of
 CC polypeptides which do not normally contain trans-4-hydroxyproline. The
 CC present sequence represents a human collagen 1 (alpha1) protein, which
 CC may be produced using the method of the invention
 XX Sequence 1057 AA;
 XX
 Query Match 97.0%; Score 1149; DB 3; Length 1057;
 Best Local Similarity 93.6%; Pred. No. 6.4e-67;
 Matches 206; Conservative 0; Mismatches 2; Indels 12; Gaps 1;
 QY 1 GPP-----GPPGPTGLPGRGSGRFPAGADGVAGPKGAGRGSPGPA 48
 DB 297 GPPGAGEGKRGAGRGFPGLGPPGGRGSGRFPAGADGVAGPKGAGRGSPGPA 356
 QY 49 GPKGSPGAGRPAGLPGAKGLTSGSPGPDGKTGPPAGQDGRPPGPPGARGQA 108
 DB 357 GPKGSPGAGRPAGLPGAKGLTSGSPGPDGKTGPPAGQDGRPPGPPGARGQA 416
 QY 109 GVMGFPKGAGRPAGRGVPGPCAVGPAKDGAGAGQGGPPGAGRGGEQGA 168
 DB 417 GVMGFPKGAGRPAGRGVPGPCAVGPAKDGAGAGQGGPPGAGRGGEQGA 476
 QY 169 GSPGFQGLPGAGPPGAGKFGEGQGVFDLGLGAPGSPGAG 208
 DB 477 GSPGFQGLPGAGPPGAGKFGEGQGVFDLGLGAPGSPGAG 516
 RESULT 5
 AAY84544
 ID AAY84544 standard; protein; 1057 AA.

XX AC AAY84544;
XX DT 25-JUL-2000 (first entry)
XX DE A human collagen 1 (alpha1) protein helical region.
XX KW Extracellular matrix protein; self aggregation; hydroxylated proline;
XX KW trans-4-hydroxyproline; 3-hydroxyproline; recombinant protein production;
XX KW collagen; fibrinogen; fibronectin; post translational hydroxylation.
XX OS Homo sapiens.
XX PN EP992586-A2.
XX PD 12-APR-2000.
XX PF 07-OCT-1999; 99EP-00119184.
XX PR 09-OCT-1998; 98US-00169768.
XX PA (USSU) US SURGICAL CORP.
XX PI Gruskin EA, Buechter DD, Zhang G, Connolly K;
XX DR WPI: 2000-259138/23.
XX DR N-PSDB; AAA12503.
XX PT Production of extracellular matrix proteins containing 4-trans-
XX PT hydroxyproline results in native self aggregating proteins, useful on
XX PT medical implants.
XX PS Example 10; Fig 39A-E; 260pp; English.
XX CC The specification describes a method for producing an extracellular
XX CC matrix protein or its fragment. The extracellular matrix protein is
XX CC capable of self aggregating in a cell which does not ordinarily
XX CC hydroxylated prolines. The method comprises optimising a nucleic acid
XX CC sequence for expression in the cell by substitution of codons preferred
XX CC by that cell for naturally occurring codons not preferred by the cell;
XX CC incorporating the nucleic acid sequence into the cell; and contacting the
XX CC cell with a hypertonic growth medium containing at least one amino acid,
XX CC selected from the group consisting of trans-4-hydroxyproline and 3-
XX CC hydroxyproline to allow at least one of the amino acids to be assimilated
XX CC into the cell and incorporated into the extracellular matrix protein. The
XX CC method may be used to make host cells assimilate and incorporate trans-4-
XX CC hydroxyproline into proteins. This is especially useful in the
XX CC recombinant production of proteins such as collagen, fibrinogen and
XX CC fibronectin whose ability to self aggregate and produce functional
XX CC proteins depends on the post translational hydroxylation of proline. The
XX CC method is also useful in studying the structure and function of
XX CC polypeptides which do not normally contain trans-4-hydroxyproline. The
XX CC present sequence represents human collagen 1 (alpha1) helical region,
XX CC which may be produced using the method of the invention
XX SQ Sequence 1057 AA;

Query Match 97.0%; Score 1149; DB 3; Length 1057;
Best Local Similarity 93.6%; Pred. No. 6.4e-67;
Matches 206; Conservative 0; Mismatches 2; Indels 12; Gaps 1;
QY 1 GPP-----GEPGPTGLPGPGERGCGSRGPGADGVAGPKGAGRGSGGPA 48
DB 297 GPPGAGEGKRGARGEPGPTGLPGPGERGCGSRGPGADGVAGPKGAGRGSGGPA 356
QY 49 GPKSGPGEAGRPGEAGLPGAKGLTSGSGSPGDKTGTGPPGAGDGRPPGPPGARGQA 108
DB 357 GPKSGPGEAGRPGEAGLPGAKGLTSGSGSPGDKTGTGPPGAGDGRPPGPPGARGQA 416
QY 109 GWMGPPGKGAAGPFGKAGRGVPGPGAVGPAKGDKGEAGQAQPPGPPGARGQA 168
DB 417 GWMGPPGKGAAGPFGKAGRGVPGPGAVGPAKGDKGEAGQAQPPGPPGARGQA 476

QY 169 GSPGQGLPGPAGPPGEGAGKPGEGQGVPGDILGAPGSPGAG 208
DB 477 GSPGQGLPGPAGPPGEGAGKPGEGQGVPGDILGAPGSPGAG 516

RESULT 6

AY84403
ID AAY84403 standard; protein; 1058 AA.
XX AC AAY84403;
XX DT 12-JUL-2000 (first entry)
XX DE Amino acid sequence of human type 1 (alpha1) collagen polypeptide.
XX KW Alpha1 collagen; 3,4-dehydro-L-proline; epoxidation; 3,4-epoxyproline;
XX KW collagen; mussel adhesive protein; bioadhesive.
XX OS Homo sapiens.
XX PN WO200014201-A1.
XX PD 16-MAR-2000.
XX PF 07-SEP-1999; 99WO-US020462.
XX PR 09-SEP-1998; 98US-0099652P.
XX PA (USSU) US SURGICAL CORP.
XX PA (PAOL/) PAOLELLA D N.
XX PA (GRUS/) GRUSKIN E A.
XX PA (BUEC/) BUECHTER D D.
XX PI Paoletta DN, Gruskin EA, Buechter DD;
XX DR WPI: 2000-271051/23.
XX DR N-PSDB; AAZ99843.
XX PT Incorporating non-natural amino acid into polypeptide, useful e.g. for
XX PT production of bioadhesives, by epoxidation or substitution of
XX PT dehydroproline residues.
XX PS Disclosure; Fig 6; 66pp; English.
XX CC The present sequence represents a human type 1 (alpha1) collagen protein.
XX CC Peptides derived from the protein were used to demonstrate incorporation
XX CC of 3,4-dehydro-L-proline into the peptide, using the method of the
XX CC invention. The specification describes a method for the incorporation of
XX CC non-natural amino acid into a polypeptide. The method comprises reacting
XX CC at least one 3,4-dehydroproline residue in the polypeptide with an
XX CC epoxidation reagent from a polypeptide containing at least one 3,4-
XX CC epoxyproline residue. The method is used for studying the effects of non-
XX CC natural amino acids on structure and function of polypeptides. The method
XX CC is also useful for commercial production of collagen or mussel adhesive
XX CC proteins (which are useful as bioadhesives), and for incorporating a wide
XX CC variety of groups, including therapeutic ligands and biological probes,
XX CC into polypeptides
XX SQ Sequence 1058 AA;

Query Match 97.0%; Score 1149; DB 3; Length 1058;
Best Local Similarity 93.6%; Pred. No. 6.4e-67;
Matches 206; Conservative 0; Mismatches 2; Indels 12; Gaps 1;
QY 1 GPP-----GEPGPTGLPGPGERGCGSRGPGADGVAGPKGAGRGSGGPA 48
DB 298 GPPGAGEGKRGARGEPGPTGLPGPGERGCGSRGPGADGVAGPKGAGRGSGGPA 357
QY 49 GPKSGPGEAGRPGEAGLPGAKGLTSGSGSPGDKTGTGPPGAGDGRPPGPPGARGQA 108
DB 358 GPKSGPGEAGRPGEAGLPGAKGLTSGSGSPGDKTGTGPPGAGDGRPPGPPGARGQA 417
QY 109 GWMGPPGKGAAGPFGKAGRGVPGPGAVGPAKGDKGEAGQAQPPGPPGARGQA 168

Matches	206;	Conservative	0;	Mismatches	2;	Indels	12;	Gaps	17.
Qy	1	GPP-----	-GEPGTGLPGPGERGGPSGRFPFGADGVAGPKGPAGERGSPGPA	48					
Dd	297	GPPGAGEGKRGARGEFPTGLTPGPPGERGGPSGRFPFGADGVAGPKGPAGERGSPGPA	356						
Qy	49	GPKGSPGEAGRPEAGLPGAAGLTGSGSPGDGKTGPPGPAGODGRPDPGPPGARQA	108						
Dd	357	GPKGSPGEAGRPEAGLPGAAGLTGSGSPGDGKTGPPGPAGODGRPDPGPPGARQA	416						
Qy	109	GVMGFFPGPKAAAGBPPGKAGRGVPGPCAVCPAGCKDGEACAGQGPPGPAGPAGERGEQCPA	168						
Dd	417	GVMGFFPGPKAAAGBPPGKAGRGVPGPCAVCPAGCKDGEACAGQGPPGPAGPAGERGEQCPA	476						
Qy	169	GSPGFQGLPGAPPPGAGPGEQGVFGDLGAPGPSGAPAG	208						
Dd	477	GSPGFQGLPGAPPPGAGPGEQGVFGDLGAPGPSGARG	516						
<hr/>									
RESULT	8								
AAY84540	ID	AAY84540 standard; protein; 1107 AA.							
XX AC	AAY84540;								
XX DT	25-JUL-2000	(first entry)							
XX XX	Amino acid sequence of a chimeric collagen 1 (alpha1)/decorin protein.								
XX DE	Extracellular matrix protein; self aggregation; hydroxylated proline;								
KW KM	trans-4-hydroxyproline; 3-hydroxyproline; recombinant protein production;								
KW KW	collagen; fibrinogen; fibronectin; post translational hydroxylation;								
KW KW	decorin; chimera.								
XX XX									
OS OS	Homo sapiens.								
OS OS	Unidentified.								
XX XX	Chimeric.								
FH FT	Key	Location/Qualifiers							
FT FT	Misc-difference	858 /note= "Gly encoded by GCT"							
XX FN	EP992586-A2.								
XX PD	12-APR-2000.								
XX PF	07-OCT-1999;	99EP-00119184.							
XX PR	09-OCT-1998;	98US-00169768.							
XX PA	(USSU) US SURGICAL CORP.								
XX PI	Gruskin EA, Buechter DD, Zhang G, Connolly K;								
XX DR	WT; 2000-259138/23.								
DR DR	N-FSD5; AAA12500.								
XX PT	Production of extracellular matrix proteins containing 4-trans-								
PT PT	hydroxyproline results in native self aggregating proteins, useful on								
PT PT	medical implants.								
XX PS	Claim 24; Fig 18; 260pp; English.								
XX CC	The specification describes a method for producing an extracellular								
CC CC	matrix protein or its fragment. The extracellular matrix protein is								
CC CC	capable of self aggregating in a cell which does not ordinarily								
CC CC	hydroxylated prolines. The method comprises optimising a nucleic acid								
CC CC	sequence for expression in the cell by substitution of codons preferred								
CC CC	by that cell for naturally occurring codons not preferred by the cell;								
CC CC	incorporating the nucleic acid sequence into the cell; and contacting the								
CC CC	cell with a hypertonic growth medium containing at least one amino acid,								
CC CC	selected from the group consisting of trans-4-hydroxyproline and 3-								
CC CC	hydroxyproline to allow at least one of the amino acids to be assimilated								

CC into the cell and incorporated into the extracellular matrix protein. The
 CC method may be used to make host cells assimilate and incorporate trans-4-
 CC hydroxyproline into proteins. This is especially useful in the
 CC recombinant production of proteins such as collagen, fibronectin and
 CC fibronectin whose ability to self aggregate and produce functional
 CC proteins depends on the post translational hydroxylation of proline. The
 CC method is also useful in studying the structure and function of
 CC polypeptides which do not normally contain trans-4-hydroxyproline. The
 CC present sequence represents a chimeric collagen 1 (alpha1)/decorin
 CC protein, which may be produced using the method of the invention
 XX
 XX Sequence 1107 AA;

Query Match 97.0%; Score 1149; DB 3; Length 1107;
 Best Local Similarity 93.6%; Pred. No. 6.6e-67;
 Matches 206; Conservative 0; Mismatches 2; Indels 12; Gaps 1;
 QY 1 GPP-----GPPGPTGLPGERGSGRPGADGVAGPKGPGAGERSGPGA 48
 DB 297 GPPPAGEEGRKARGEPGPTGLPGERGSGRPGADGVAGPKGPGAGERSGPGA 356
 QY 49 GPKSGPGEAGRPGEAGLPGAKLTGSPGSPGDKTGTGPPAGQDGRPPGPPGARGQA 108
 DB 357 GPKSGPGEAGRPGEAGLPGAKLTGSPGSPGDKTGTGPPAGQDGRPPGPPGARGQA 416
 QY 109 GVMGFPKGAAGSPKAGRGVPGPCAVGPKAGKDGAGAGQGPAGPAGERSGPGA 168
 DB 417 GVMGFPKGAAGSPKAGRGVPGPCAVGPKAGKDGAGAGQGPAGPAGERSGPGA 476
 QY 169 GSPFGQGLPGPAGPGEAGKPGEGQVPGDLGAPGSPGAG 208
 DB 477 GSPFGQGLPGPAGPGEAGKPGEGQVPGDLGAPGSPGAG 516

RESULT 9
 ADE87050
 ID ADE87050 standard; protein; 1161 AA.
 XX
 AC ADE87050;
 XX
 DT 29-JAN-2004 (first entry)
 XX
 DE Human pancreatic cell protein sequence SegID510.
 XX
 KW neoplastic pancreatic cell; pancreatic cancer;
 KW cancer death; cytostatic; vaccine; gene therapy;
 KW non-cancerous pancreas disease; human.
 XX
 OS Homo sapiens.
 XX
 PN WO2003060145-A2.
 XX
 PD 24-JUL-2003.
 XX
 PF 19-DEC-2002; 2002WO-US040655.
 XX
 PR 21-DEC-2001; 2001US-0342768P.
 XX
 PA (DIAD-) DIADEXUS INC.
 XX
 PI Sun Y, Liu C;
 XX
 DR WPI; 2003-587286/55.
 DR N-PSDB; ADE87387.
 XX
 PT New pancreatic specific nucleic acid molecule or protein for diagnosing,
 PT staging, imaging, monitoring, preventing or treating pancreatic cancer or
 PT non-cancerous disease states of the pancreas.
 XX
 PS Claim 12; SEQ ID NO 510; 635pp; English.
 CC This invention relates to novel nucleic acids and proteins present in
 CC normal and neoplastic pancreatic cells. Pancreatic cancer is a common

CC cause of cancer death worldwide, therefore accurate methods of diagnosis
 CC and treatment are required. Compounds which modulate the proteins of the
 CC invention may have cytostatic activity and the protein and DNA sequences
 CC of the invention may be useful for the development of a vaccine or in
 CC gene therapy. The composition and methods are useful in diagnosing,
 CC staging, imaging, monitoring, preventing or treating pancreatic cancer
 CC and non-cancerous disease states of the pancreas. The present sequence is
 CC that of a human pancreatic protein of the invention.
 XX
 XX Sequence 1161 AA;

Query Match 97.0%; Score 1149; DB 7; Length 1161;
 Best Local Similarity 93.6%; Pred. No. 6.9e-67;
 Matches 206; Conservative 0; Mismatches 2; Indels 12; Gaps 1;
 QY 1 GPP-----GPPGPTGLPGERGSGRPGADGVAGPKGPGAGERSGPGA 48
 DB 155 GPPPAGEEGRKARGEPGPTGLPGERGSGRPGADGVAGPKGPGAGERSGPGA 214
 QY 49 GPKSGPGEAGRPGEAGLPGAKLTGSPGSPGDKTGTGPPAGQDGRPPGPPGARGQA 108
 DB 215 GPKSGPGEAGRPGEAGLPGAKLTGSPGSPGDKTGTGPPAGQDGRPPGPPGARGQA 274
 QY 109 GVMGFPKGAAGSPKAGRGVPGPCAVGPKAGKDGAGAGQGPAGPAGERSGPGA 168
 DB 275 GVMGFPKGAAGSPKAGRGVPGPCAVGPKAGKDGAGAGQGPAGPAGERSGPGA 334
 QY 169 GSPFGQGLPGPAGPGEAGKPGEGQVPGDLGAPGSPGAG 208
 DB 335 GSPFGQGLPGPAGPGEAGKPGEGQVPGDLGAPGSPGAG 374

RESULT 10
 AAR89469
 ID AAR89469 standard; protein; 1169 AA.
 XX
 AC AAR89469;
 XX
 DT 01-OCT-1996 (first entry)
 XX
 DE Collagen/BMP-2B fusion protein.
 XX
 KW Bone morphogenic protein 2B; BMP-2B; collagen IA; osteogenesis;
 KW fusion protein.
 XX
 OS Synthetic.
 XX
 PH Key
 FT Domain
 FT 1..1057
 FT /label= Collagen-IA
 FT /note= "collagen IA alpha-helical domain"
 FT Misc-difference 887
 FT /note= "unidentified amino acid"
 FT Misc-difference 890
 FT /note= "unidentified amino acid"
 FT Peptide
 FT 1058..1059
 FT /label= Linker_peptide
 FT Domain
 FT 1060..1169
 FT /label= BMP-2B
 FT /note= "human mature BMP-2B"
 XX
 CA2151547-A.
 XX
 PD 11-DEC-1995.
 XX
 PF 12-JUN-1995; 95CA-02151547.
 XX
 PR 10-JUN-1994; 94US-00259263.
 XX
 PA (USSU) US SURGICAL CORP.
 XX
 PI Gruskin EA, Espino P;
 XX

OS Synthetic.
 XX Key
 FH Domain
 FT Location/Qualifiers
 FT 1. 1057
 FT /label= "collagen-IA
 FT /note= "collagen IA alpha-helical domain"
 FT Misc-difference 887
 FT /note= "unidentified amino acid"
 FT Misc-difference 890
 FT /note= "unidentified amino acid"
 FT Peptide
 FT 1058. 1059
 FT /label= Linker_peptide
 FT Domain
 FT 1060. 1171
 FT /label= TGF-beta-1
 FT /note= "human mature TGF-beta-1"
 XX CA2151547-A.
 PN 11-DEC-1995.
 PD 12-JUN-1995; 95CA-02151547.
 XX 10-JUN-1994; 94US-00259263.
 XX (USU) US SURGICAL CORP.
 PA Gruskin EA, Espino P;
 PI WPI; 1996-140144/15.
 DR N-PSDB; AAT16516.
 XX Chimaeric DNA encoding protein contg. extracellular matrix protein domain
 PT - and cellular regulatory factor domain, partic. useful as osteogenic
 PT agents, also related vectors, transformed cells and chimaeric proteins.
 XX Disclosure; Fig 6; 59pp; English.
 PS A fusion protein (AAR89470) comprises the alpha-helical region of human
 CC collagen I(a) linked to the human mature transforming growth factor beta-
 CC 1 (TGF-beta-1). It can be expressed in Escherichia coli transformants
 CC carrying a vector incorporating a chimeric gene (AAT16516) coding for the
 CC fusion. The TGF-beta- moiety increases efficacy of the body's normal soft
 CC tissue repair response and also induces osteogenesis. The collagen moiety
 CC provides an integral substratum or scaffolding for the TGF and cells
 CC involved in reconstruction and growth. The fusion protein provides
 CC sustained release and delivery of TGF-beta-1 to a target tissue
 XX Sequence 1171 AA;
 SQ Query Match 97.0%; Score 1149; DB 2; Length 1171;
 Best Local Similarity 93.6%; Pred. No. 7e-67;
 Matches 206; Conservative 0; Mismatches 2; Indels 12; Gaps 1;
 QY 1 GPP-----GPPGPTGLPGRGSGRFGPCADGVAGPKGPGAGRGSGGPA 48
 Db 297 GPPGAGEGKRGARGPPTGLPGRGSGRFGPCADGVAGPKGPGAGRGSGGPA 356
 QY 49 GPKGSPGEAGRPGAGLPGAKGLTSGSPGPDGKTGPPGAGQDGRPPGPPGARGQA 108
 Db 357 GPKGSPGEAGRPGAGLPGAKGLTSGSPGPDGKTGPPGAGQDGRPPGPPGARGQA 416
 QY 109 GVMGFPKPGKAGPFGKAGRGVPPGAVGPPAGKDGCAQAQPPGPPGAGRGQGGA 168
 Db 417 GVMGFPKPGKAGPFGKAGRGVPPGAVGPPAGKDGCAQAQPPGPPGAGRGQGGA 476
 QY 169 GSPGFQGLPAGPPGAGPFGEGVPGDGLGARGPSGPA 208
 Db 477 GSPGFQGLPAGPPGAGPFGEGVPGDGLGARGPSGARG 516

RESULT 13
 AAY84538

ID AAY84538 standard; protein; 1171 AA.
 XX
 AC AAY84538;
 XX
 DT 25-JUL-2000 (first entry)
 XX
 DE A chimeric collagen 1 (alpha1)/TGF-beta1 protein.
 XX
 KW Extracellular matrix protein; self aggregation; hydroxylated proline;
 KW trans-4-hydroxyproline; 3-hydroxyproline; recombinant protein production;
 KW collagen; fibronectin; fibronectin; post translational hydroxylation;
 KW ss. transforming growth factor-beta1; TGF-beta1; chimera.
 XX
 OS Homo sapiens.
 OS Unidentified.
 OS Chimeric.
 XX
 FH Key Location/Qualifiers
 FT Misc-difference 858
 FT /note= "Gly encoded by GCT"
 FT
 FT
 FT
 FN EP992586-A2.
 XX
 PD 12-APR-2000.
 XX
 XX 07-OCT-1999; 99EP-00119184.
 PF
 XX 09-OCT-1998; 98US-00169768.
 PR
 XX (USU) US SURGICAL CORP.
 PA Gruskin EA, Buechter DD, Zhang G, Connolly K;
 XX WPI; 2000-259138/23.
 DR N-PSDB; AAA12498.
 XX
 PT Production of extracellular matrix proteins containing 4-trans-
 PT hydroxyproline results in native self aggregating proteins, useful on
 PT medical implants.
 XX
 XX Claim 23; Fig 15; 260pp; English.
 PS The specification describes a method for producing an extracellular
 CC matrix protein or its fragment. The extracellular matrix protein is
 CC capable of self aggregating in a cell which does not ordinarily
 CC hydroxylated prolines. The method comprises optimising a nucleic acid
 CC sequence for expression in the cell by substitution of codons preferred
 CC by that cell for naturally occurring codons not preferred by the cell;
 CC incorporating the nucleic acid sequence into the cell; and contacting the
 CC cell with a hypertonic growth medium containing at least one amino acid,
 CC selected from the group consisting of trans-4-hydroxyproline and 3-
 CC hydroxyproline to allow at least one of the amino acids to be assimilated
 CC into the cell and incorporated into the extracellular matrix protein. The
 CC method may be used to make host cells assimilate and incorporate trans-4-
 CC hydroxyproline into proteins. This is especially useful in the
 CC recombinant production of proteins such as collagen, fibrinogen and
 CC fibronectin whose ability to self aggregate and produce functional
 CC proteins depends on the post translational hydroxylation of proline. The
 CC method is also useful in studying the structure and function of
 CC polypeptides which do not normally contain trans-4-hydroxyproline. The
 CC present sequence represents chimeric collagen 1 (alpha1)/transforming
 CC growth factor-beta1 (TGF-beta1) protein, which may be produced using the
 CC method of the invention
 XX
 SQ Sequence 1171 AA;
 Query Match 97.0%; Score 1149; DB 3; Length 1171;
 Best Local Similarity 93.6%; Pred. No. 7e-67;
 Matches 206; Conservative 0; Mismatches 12; Gaps 1;
 QY 1 GPP-----GPPGPTGLPGRGSGRFGPCADGVAGPKGPGAGRGSGGPA 48
 Db 297 GPPGAGEGKRGARGPPTGLPGRGSGRFGPCADGVAGPKGPGAGRGSGGPA 356

QY 49 GPKSGPGEAGRPGEAGLPGCAKGLTSGPSGPDGKTPPGPAGDGRPPGPPGARGQA 108
DB 357 GPKSGPGEAGRPGEAGLPGCAKGLTSGPSGPDGKTPPGPAGDGRPPGPPGARGQA 416
QY 109 GVMGFPKGAAGBPGKAGRGVPGPCAVCPAGKDGAGAGAGQPPGAGPAGERGEGGPA 168
DB 417 GVMGFPKGAAGBPGKAGRGVPGPCAVCPAGKDGAGAGAGQPPGAGPAGERGEGGPA 476
QY 169 GSPGFQGLPGPAGPPGAGKPGEGQVPGDGLGAPGSPGAG 208
DB 477 GSPGFQGLPGPAGPPGAGKPGEGQVPGDGLGAPGSPGARG 516

RESULT 14
ADE87057
ID ADE87057 standard; protein; 1211 AA.

XX AC ADE87057;
XX DT 29-JAN-2004 (first entry)
XX DE Human pancreatic cell protein sequence SeqID517.
XX KW neoplastic pancreatic cell; pancreatic cancer;
XX KW cancer death; cytostatic; vaccine; gene therapy;
XX KW non-cancerous pancreas disease; human.
XX OS Homo sapiens.
XX PN WO2003060145-A2.
XX PD 24-JUL-2003.
XX PF 19-DEC-2002; 2002WO-US040655.
XX PR 21-DEC-2001; 2001US-0342768P.
XX PA (DIAD-) DIADEXUS INC.

XX PI Sun Y, Liu C;
XX DR WPI; 2003-587286/55.
XX DR N-PSDB; ADE87397.
XX PT New pancreatic specific nucleic acid molecule or protein for diagnosing,
PT staging, imaging, monitoring, preventing or treating pancreatic cancer or
PT non-cancerous disease states of the pancreas.
XX PS Claim 12; SEQ ID NO 517; 635pp; English.

XX CC This invention relates to novel nucleic acids and proteins present in
CC normal and neoplastic pancreatic cells. Pancreatic cancer is a common
CC cause of cancer death worldwide, therefore accurate methods of diagnosis
CC and treatment are required. Compounds which modulate the proteins of the
CC invention may have cytostatic activity and the protein and DNA sequences
CC of the invention may be useful for the development of a vaccine or in
CC gene therapy. The composition and methods are useful in diagnosing,
CC staging, imaging, monitoring, preventing or treating pancreatic cancer is
CC and non-cancerous disease states of the pancreas. The present sequence is
CC that of a human pancreatic protein of the invention.

XX SQ Sequence 1211 AA;
Query Match 97.0%; Score 1149; DB 7; Length 1211;
Best Local Similarity 93.6%; Pred. No. 7.2e-67;
Matches 206; Conservative 0; Mismatches 2; Indels 12; Gaps 1;

QY 1 GPP-----GEPGPTGLPGPPGGRGGSGRGGAGDVAGPKGPPAGERGSPGA 48
DB 458 GPPGPAEGEGRGARGEGPTGLPGPPGGRGGSGRGGAGDVAGPKGPPAGERGSPGA 517
QY 49 GPKSGPGEAGRPGEAGLPGCAKGLTSGPSGPDGKTPPGPAGDGRPPGPPGARGQA 108

DB 518 GPKSGPGEAGRPGEAGLPGCAKGLTSGPSGPDGKTPPGPAGDGRPPGPPGARGQA 577
QY 109 GVMGFPKGAAGBPGKAGRGVPGPCAVCPAGKDGAGAGAGQPPGAGPAGERGEGGPA 168
DB 578 GVMGFPKGAAGBPGKAGRGVPGPCAVCPAGKDGAGAGAGQPPGAGPAGERGEGGPA 637
QY 169 GSPGFQGLPGPAGPPGAGKPGEGQVPGDGLGAPGSPGAG 208
DB 638 GSPGFQGLPGPAGPPGAGKPGEGQVPGDGLGAPGSPGARG 677

RESULT 15
ADE87062
ID ADE87062 standard; protein; 1226 AA.

XX AC ADE87062;
XX DT 29-JAN-2004 (first entry)
XX DE Human pancreatic cell protein sequence SeqID522.
XX KW neoplastic pancreatic cell; pancreatic cancer;
XX KW cancer death; cytostatic; vaccine; gene therapy;
XX KW non-cancerous pancreas disease; human.
XX OS Homo sapiens.
XX PN WO2003060145-A2.
XX PD 24-JUL-2003.
XX PF 19-DEC-2002; 2002WO-US040655.
XX PR 21-DEC-2001; 2001US-0342768P.
XX PA (DIAD-) DIADEXUS INC.

XX PI Sun Y, Liu C;
XX DR WPI; 2003-587286/55.
XX DR N-PSDB; ADE87403.
XX PT New pancreatic specific nucleic acid molecule or protein for diagnosing,
PT staging, imaging, monitoring, preventing or treating pancreatic cancer or
PT non-cancerous disease states of the pancreas.

XX PS Claim 12; SEQ ID NO 522; 635pp; English.
XX CC This invention relates to novel nucleic acids and proteins present in
CC normal and neoplastic pancreatic cells. Pancreatic cancer is a common
CC cause of cancer death worldwide, therefore accurate methods of diagnosis
CC and treatment are required. Compounds which modulate the proteins of the
CC invention may have cytostatic activity and the protein and DNA sequences
CC of the invention may be useful for the development of a vaccine or in
CC gene therapy. The composition and methods are useful in diagnosing,
CC staging, imaging, monitoring, preventing or treating pancreatic cancer is
CC and non-cancerous disease states of the pancreas. The present sequence is
CC that of a human pancreatic protein of the invention.

XX SQ Sequence 1226 AA;
Query Match 97.0%; Score 1149; DB 7; Length 1226;
Best Local Similarity 93.6%; Pred. No. 7.2e-67;
Matches 206; Conservative 0; Mismatches 2; Indels 12; Gaps 1;

QY 1 GPP-----GEPGPTGLPGPPGGRGGSGRGGAGDVAGPKGPPAGERGSPGA 48
DB 458 GPPGPAEGEGRGARGEGPTGLPGPPGGRGGSGRGGAGDVAGPKGPPAGERGSPGA 517
QY 49 GPKSGPGEAGRPGEAGLPGCAKGLTSGPSGPDGKTPPGPAGDGRPPGPPGARGQA 108
DB 518 GPKSGPGEAGRPGEAGLPGCAKGLTSGPSGPDGKTPPGPAGDGRPPGPPGARGQA 577

109	QY	GVMGFPFGKGAAGSPGKAGRGVFPBPFGAVGPAGKDGEAGAOGPPGPGPAGPAGERGEOGPA	168
578	DB	GVMGFPFGKGAAGSPGKAGRGVFPBPFGAVGPAGKDGEAGAOGPPGPGPAGPAGERGEOGPA	637
169	QY	GSPGFQGLPGPAGPAPGGAGXPGQGVPGDLAGFPGSPGAG	208
638	DB	GSPGFQGLPGPAGPAPGGAGXPGQGVPGDLAGFPGSPGAG	577

Search completed: September 24, 2004, 11:09:14
Job time : 31.5883 secs

Sequence 1, Appl
Sequence 1, Appl
Sequence 79, Appl
Sequence 2, Appl
Sequence 2, Appl
Sequence 49, Appl
Sequence 26, Appl
Sequence 8, Appl
Sequence 8, Appl
Sequence 18, Appl
Sequence 10, Appl
Sequence 1, Appl
Sequence 20, Appl
Sequence 5, Appl
Sequence 48, Appl
Sequence 48, Appl
Sequence 50, Appl
Sequence 22, Appl
Sequence 33, Appl
Sequence 12, Appl
Sequence 10, Appl
Sequence 10, Appl
Sequence 70, Appl
Sequence 35, Appl
Sequence 74, Appl
Sequence 22, Appl
Sequence 23, Appl
Sequence 4, Appl

16 2966 96.6 1464 16 US-10-357-851-1
17 2966 96.6 1464 16 US-10-358-024-1
18 2966 96.6 1464 16 US-10-734-564-79
19 2936 95.6 1463 15 US-10-402-089-2
20 2936 95.6 1463 15 US-10-402-072A-2
21 2857 93.1 822 15 US-10-342-331-49
22 2857 93.1 1453 16 US-10-488-091-26
23 2816.5 91.7 1449 15 US-10-402-089-8
24 2816.5 91.7 1449 15 US-10-402-072A-8
25 2791.5 90.9 1341 14 US-10-058-124-18
26 2226 72.5 1017 16 US-10-639-286-10
27 2209 72.0 1014 14 US-10-194-441A-1
28 2202 71.7 1418 14 US-10-058-124-20
29 2202 71.7 1418 16 US-10-488-091-5
30 2189 71.3 1014 14 US-10-194-441A-48
31 2189 71.3 1487 16 US-10-488-091-6
32 2156 70.2 595 15 US-10-342-331-48
33 2156 70.2 595 15 US-10-342-331-50
34 2045 66.6 500 12 US-10-232-175-22
35 2045 66.6 662 12 US-10-232-175-33
36 1924.5 62.7 1466 15 US-10-402-089-12
37 1924.5 62.7 1466 15 US-10-402-072A-12
38 1923 62.6 1366 15 US-10-402-089-10
39 1923 62.6 1366 15 US-10-402-072A-10
40 1921.5 62.6 1496 14 US-10-177-293-70
41 1921.5 62.6 1496 14 US-10-301-822-35
42 1921.5 62.6 1496 15 US-10-236-031B-74
43 1921.5 62.6 1496 16 US-10-488-091-22
44 1910 62.2 1497 16 US-10-488-091-23
45 1906.5 62.1 1466 15 US-10-402-089-4

ALIGNMENTS

RESULT 1
US-10-104-889-16
; Sequence 16, Application US/10104889
; Publication No. US20040086961A1
; GENERAL INFORMATION:
; APPLICANT: GRUSKIN, ELLIOT A.
; BUECHTER, DOUGLAS
; BROKAW, JANE
; ZHANG, GUANGHUI
; PAOLELLA, DAVID
; TITLE OF INVENTION: AMINO ACID MODIFIED POLYPEPTIDES
; NUMBER OF SEQUENCES: 50
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: DILWORTH & BARRESE
; STREET: 333 EARLE OVINGTON BOULEVARD
; CITY: UNIONDALE
; STATE: NY
; COUNTRY: U.S.A.
; ZIP: 11553
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION NUMBER: US/10/104,889
; FILING DATE: 22-Mar-2002
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/09/169,768
; FILING DATE: 09-OCT-1998
; ATTORNEY/AGENT INFORMATION:
; NAME: STEEN, JEFFREY S
; TELEPHONE: (516) 228-8484
; TELEFAX: (516) 228-8516
; INFORMATION FOR SEQ ID NO: 16:
; SEQUENCE CHARACTERISTICS:

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: September 24, 2004, 11:07:31 ; Search time 67.1873 Seconds
(without alignments)
2603.588 Million cell updates/sec

Title: US-10-658-989A-4

Perfect score: 3070

Sequence: 1 GSEGPVGRGPPGPPGAGA.....PGPSGDAGPPGPPGPKGEG 544

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1349238 seqs, 32158718 residues

Total number of hits satisfying chosen parameters: 1349238

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Listing first 45 summaries

Database : Published Applications AA:*

- 1: /cgn2_6/ptodata/1/pubpaa/US07_PUBCOMB.pep.*
- 2: /cgn2_6/ptodata/1/pubpaa/PCT_NEW_PUB.pep.*
- 3: /cgn2_6/ptodata/1/pubpaa/US06_NEW_PUB.pep.*
- 4: /cgn2_6/ptodata/1/pubpaa/US06_PUBCOMB.pep.*
- 5: /cgn2_6/ptodata/1/pubpaa/US07_NEW_PUB.pep.*
- 6: /cgn2_6/ptodata/1/pubpaa/PCTUS_PUBCOMB.pep.*
- 7: /cgn2_6/ptodata/1/pubpaa/US08_NEW_PUB.pep.*
- 8: /cgn2_6/ptodata/1/pubpaa/US08_PUBCOMB.pep.*
- 9: /cgn2_6/ptodata/1/pubpaa/US09A_PUBCOMB.pep.*
- 10: /cgn2_6/ptodata/1/pubpaa/US09B_PUBCOMB.pep.*
- 11: /cgn2_6/ptodata/1/pubpaa/US09C_PUBCOMB.pep.*
- 12: /cgn2_6/ptodata/1/pubpaa/US09_NEW_PUB.pep.*
- 13: /cgn2_6/ptodata/1/pubpaa/US10A_PUBCOMB.pep.*
- 14: /cgn2_6/ptodata/1/pubpaa/US10B_PUBCOMB.pep.*
- 15: /cgn2_6/ptodata/1/pubpaa/US10C_PUBCOMB.pep.*
- 16: /cgn2_6/ptodata/1/pubpaa/US10_NEW_PUB.pep.*
- 17: /cgn2_6/ptodata/1/pubpaa/US60_NEW_PUB.pep.*
- 18: /cgn2_6/ptodata/1/pubpaa/US60_PUBCOMB.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2966	96.6	1057	16	US-10-104-889-16
2	2966	96.6	1057	16	US-10-104-889-20
3	2966	96.6	1107	16	US-10-104-889-11
4	2966	96.6	1169	16	US-10-104-889-6
5	2866	96.6	1171	16	US-10-104-889-8
6	2966	96.6	1388	16	US-10-104-889-10
7	2966	96.6	1461	16	US-10-488-091-25
8	2966	96.6	1464	12	US-09-918-715-261
9	2966	96.6	1464	14	US-10-060-036-159
10	2966	96.6	1464	14	US-10-171-311-36
11	2966	96.6	1464	14	US-10-216-705-21
12	2966	96.6	1464	14	US-10-149-352-2
13	2966	96.6	1464	14	US-10-177-293-65
14	2966	96.6	1464	14	US-10-301-822-28
15	2966	96.6	1464	15	US-10-291-265-243

[illegible]


```
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
  APPLICATION NUMBER: US/10/104,889
  FILING DATE: 22-Mar-2002
  CLASSIFICATION: <Unknown>
PRIOR APPLICATION DATA:
  APPLICATION NUMBER: US/09/169,768
  FILING DATE: 09-OCT-1998
ATTORNEY/AGENT INFORMATION:
  NAME: STEEN, JEFFREY S
  TELECOMMUNICATION INFORMATION:
    TELEPHONE: (516) 228-8484
    TELEFAX: (516) 228-8516
INFORMATION FOR SEQ ID NO: 10:
  SEQUENCE CHARACTERISTICS:
    LENGTH: 1388 amino acids
    TYPE: amino acid
    STRANDEDNESS: single
    TOPOLOGY: unknown
  MOLECULE TYPE: peptide
  SEQUENCE DESCRIPTION: SEQ ID NO: 10:
US-10-104-889-10

Query Match      96.6%; Score 2966; DB 16; Length 1388;
Best Local Similarity 95.6%; Pred. No. 6.1e-146;
Matches 520; Conservative 21; Mismatches 3; Indels 0; Gaps 0;

QY 1 GSEGPVGVRGPPGPPGAGAGPAGDPCGADGEPGAKGADGAGPAGTGTAGAGPAGRGSGPGE 60
DB 201 GSEGPVGVRGPPGPPGAGAGPAGDPCGADGEPGAKGADGAGPAGTGTAGAGPAGRGSGPGE 260
QY 61 GPGGPPGPKGDSGPPGAPGAGPAGDPCGADGEPGAKGADGAGPAGTGTAGAGPAGRGSGPGE 120
DB 261 GPGGPPGPKGNSGPPGAPGAGPAGDPCGADGEPGAKGADGAGPAGTGTAGAGPAGRGSGPGE 320
QY 121 GPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPG 180
DB 321 GPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPG 380
QY 181 GSPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPG 240
DB 381 GSPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPG 440
QY 241 GPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPG 300
DB 441 GPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPG 500
QY 301 GVPDGLGAPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPG 360
DB 501 GVPDGLGAPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPG 560
QY 361 GAPGLEMPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPG 420
DB 561 GAPGLEMPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPG 620
QY 421 GESGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPG 480
DB 621 GESGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPG 680
QY 481 GPAGPAGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPG 540
DB 681 GPAGPAGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPG 740
QY 541 GKEG 544
DB 741 GKEG 744

RESULT 7
US-10-468-091-25
; Sequence 25, Application US/10468091

COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
  APPLICATION NUMBER: US/10/104,899
  FILING DATE: 22-Mar-2002
  CLASSIFICATION: <Unknown>
PRIOR APPLICATION DATA:
  APPLICATION NUMBER: US/09/169,768
  FILING DATE: 09-OCT-1998
ATTORNEY/AGENT INFORMATION:
  NAME: STEEN, JEFFREY S
  TELECOMMUNICATION INFORMATION:
    TELEPHONE: (516) 228-8484
    TELEFAX: (516) 228-8516
INFORMATION FOR SEQ ID NO: 10:
  SEQUENCE CHARACTERISTICS:
    LENGTH: 1388 amino acids
    TYPE: amino acid
    STRANDEDNESS: single
    TOPOLOGY: unknown
  MOLECULE TYPE: peptide
  SEQUENCE DESCRIPTION: SEQ ID NO: 10:
US-10-104-889-10

Query Match      96.6%; Score 2966; DB 16; Length 1388;
Best Local Similarity 95.6%; Pred. No. 6.1e-146;
Matches 520; Conservative 21; Mismatches 3; Indels 0; Gaps 0;

QY 1 GSEGPVGVRGPPGPPGAGAGPAGDPCGADGEPGAKGADGAGPAGTGTAGAGPAGRGSGPGE 60
DB 201 GSEGPVGVRGPPGPPGAGAGPAGDPCGADGEPGAKGADGAGPAGTGTAGAGPAGRGSGPGE 260
QY 61 GPGGPPGPKGDSGPPGAPGAGPAGDPCGADGEPGAKGADGAGPAGTGTAGAGPAGRGSGPGE 120
DB 261 GPGGPPGPKGNSGPPGAPGAGPAGDPCGADGEPGAKGADGAGPAGTGTAGAGPAGRGSGPGE 320
QY 121 GPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPG 180
DB 321 GPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPG 380
QY 181 GSPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPG 240
DB 381 GSPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPG 440
QY 241 GPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPG 300
DB 441 GPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPG 500
QY 301 GVPDGLGAPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPG 360
DB 501 GVPDGLGAPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPG 560
QY 361 GAPGLEMPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPG 420
DB 561 GAPGLEMPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPG 620
QY 421 GESGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPG 480
DB 621 GESGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPG 680
QY 481 GPAGPAGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPG 540
DB 681 GPAGPAGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPG 740
QY 541 GKEG 544
DB 741 GKEG 744

RESULT 8
US-09-918-715-261
; Sequence 261, Application US/09918715
; Publication No. US20030017157A1
; GENERAL INFORMATION:
; APPLICANT: Brad St. Croix
; APPLICANT: Bert Vogelstein
; APPLICANT: Kenneth Kinzler
; TITLE OF INVENTION: ENDOTHELIAL CELL EXPRESSION PATTERNS
; FILE REFERENCE: 1107.00134
```



```

; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-149-352-2

Query Match          96.6%; Score 2966; DB 14; Length 1464;
Best Local Similarity 95.6%; Pred. No. 6, 4e-146;
Matches 520; Conservative 21; Mismatches 3; Indels 0; Gaps 0

QY      1  GSEBEGVRGEGPGRPGAGAGAGADPGADGEPGCAKAGADGAPGIAGAPGPPGARGSPGE 60
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db      362 GSEPGGVGREGPGRPGAGAGAPGNADGQCPKAGANGAPGIAGAPGPPGARGSPGQ 421
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY      61  GPGCGPPRGDGSBGPAGSGKDTGAKGEPGVGVEGPPGAGEEGKPGARGBPGTGLP 120
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db      422 GCGCGPPKGNSEFOAGSGSKDGTAKGEPGVGVQPPGPAEEGKRGARGBPGTGLP 481
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY      121 GPPBERGGPSRGFFPADGVAGPKPAGERGSPGAPKSGPGEAGRCPEAGLPGAKGLT 180
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db      482 GPPBERGGPSRGFFPADGVAGPKPAGERGSPGAPKSGPGEAGRCPEAGLPGAKGLT 541
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY      181 GSPGSPGPDGKTGPPGAPAGEDRCPPGPPGARGCEAGVMPGPKNACEPGKAGRGVP 240
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db      542 GSPGSPGPDGKTGPPGAPAGEDRCPPGPPGARGQAGVMPGPKGAAGEPGKAGRGVP 601
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY      241 GPPGAVGPAKDGGEAGAGPFPGAPGAGERGEEGPAGSPGFEGLPGAPPPGAEAGKPGSE 300
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db      602 GPPGAVGPAKDGGEAGAGPFPGAPGAGERGEEGQSPAGSPGFQGLPGAPPPGCEANGKPGEQ 661
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY      301 GVPDGLGAPSPGARGEPGFGRGVRGVEGPPGADGAPDGDGAKGDGAGAPGSE 360
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db      662 GVPDGLGAPSPGARGERGFGRGVRGVEGPPGAPRGANGANGAKGDGAGAPGSPQ 721
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY      361 GAPGLEGNMGERGAAGLPGPKDRGDGAPKGDGSPKDGVRGLTGPFGPPGAPAGPDK 420
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db      722 GAPGLQWPFGERGAAGLPGPKDRGDGAPKGDGSPKDGVRGLTGPFGPPGAPGAPDX 781
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY      421 GSGSPSGPAGPTGARGAPDGRGEPGPPGAFGAPFGPADGEPGAKGEPDGAAGDAGPP 480
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db      782 GESGSPSGPAGPTGARGAPDGRGEPGPPGAPFGAPFGPADGQCPKAGKEPDAKAGDAGPP 841
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY      481 GPAGPAGPPGPTGDYGAPGARGSGSAGPCCATGFCGAAAGRVGPPSGDAGPPGCPGA 540
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db      842 GPAGPAGPPGPTGNVCAPCAKARGSAGPCCATGFCGAAAGRVGPPSGNAGPPGPPGPA 901
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY      541 GXEK 544
      : : : : :
Db      902 GXEK 905
      : : : : :

```

RESULT 13
US-10-177-293-65
; Sequence 65, Application US/10177293
; Publication No. US20030124128A1
; GENERAL INFORMATION:
; APPLICANT: Lillie, James
; APPLICANT: Glatt, Karen
; APPLICANT: Zhao, Xumei
; APPLICANT: Cannavaru, Manjula
; APPLICANT: Kamatkar, Shubhangi
; APPLICANT: Mertens, Maureen
; APPLICANT: Myer, Vic
; APPLICANT: Wang, Youzhen
; APPLICANT: Xu, Yongsiao
; APPLICANT: Hoersch, Sebastian
; APPLICANT: Monahan, John
; APPLICANT: Meyers, Rachel E.
; APPLICANT: East Jr., Robert C.
; APPLICANT: Hortobagyi, Gabriel N.
; APPLICANT: Pusztai Lajos
; APPLICANT: Meric, Funda
; APPLICANT: Sahin, Aysegul
; APPLICANT: Mills, Gordon B.
TITLE OF INVENTION: COMPOSITIONS, KITS, AND METHODS FOR IDENTIFICATION, ASSESSMENT,

```

; TITLE OF INVENTION: PREVENTION, AND THERAPY OF BREAST CANCER
; FILE REFERENCE: MRI-038
; CURRENT APPLICATION NUMBER: US/10/177,293
; CURRENT FILING DATE: 2002-06-21
; PRIOR APPLICATION NUMBER: US 60/299,887
; PRIOR FILING DATE: 2001-06-21
; PRIOR APPLICATION NUMBER: US 60/301,572
; PRIOR FILING DATE: 2001-06-27
; PRIOR APPLICATION NUMBER: US 60/306,501
; PRIOR FILING DATE: 2001-07-18
; PRIOR APPLICATION NUMBER: US 60/325,002
; PRIOR FILING DATE: 2001-09-25
; PRIOR APPLICATION NUMBER: US 60/362,585
; PRIOR FILING DATE: 2002-03-05
; PRIOR APPLICATION NUMBER: US 60/xxx,xxx
; PRIOR FILING DATE: 2002-05-14
; NUMBER OF SEQ ID NOS: 506
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 65
; LENGTH: 1464
; TYPE: PRT
; ORGANISM: Homo sapiens
; US-10-177-293-65

```

Query Match	96.6%	Score 2966;	DB 14;	Length 1464;
Best Local Similarity	95.6%;	Pred. No. 6.4e-146;		
Matches 520;	Conservative 21;	Mismatches 3;	Indels 0;	Gaps 0;
QY	1	GSEGGEGVRGPPGPPGAGAAAGPAGDPCADCEPCAKGADCAPGIACAGPFPAGRGSPGE	60	
DB	362	GSEGGEGVRGPPGPPGAGAAAGPAGDPCADQPCAKGANGAPGIACAGPFPAGRGSPGQ	421	
QY	61	GPCGPPGKGSPPGAPGSGKDTGAKGEPVGVGVEPPGPAEEGKPGARGEPGTGILP	120	
DB	422	GGCGPPGPKGNSPPGAPGSGKDTGAKGEPVGVGVPPGPAEEGKRGARGEPGPTGILP	481	
QY	121	GPFGRRGPGSRGPPGADGVAGPKPAGERGSPGAPGKSGPGEAGRPGEAGLPGAKGLT	180	
DB	482	GPFGRRGPGSRGPPGADGVAGPKPAGERGSPGAPGKSGPGEAGRPGEAGLPGAKGLT	541	
QY	181	GPSGSPGPDGKTGPPGPAGEGDRGPPGPPCARGEAAGVGFPGPKAAAGEPCKAGRGVP	240	
DB	542	GPSGSPGPDGKTGPPGPAGQDRGPPGPPCARGQAQAGVGFPGPKAAAGEPCKAGRGVP	601	
QY	241	GPFGVAVGAGKDGAGAEAGPPGPAGPAGERGEGPAGSPGFEGLPGPAGPGEAGKPGSE	300	
DB	602	GPFGVAVGAGKDGAGAGPPGPAGPAGERGEGQAGSPGFQGLPGPAGPGEAGKPGSQ	661	
QY	301	GVPGDLGAPGFSGARGPFPFPGRGVSGPPGPAGPPGADGAPDCAKGDAGACAPAGSE	360	
DB	662	GVPGDLGAPGFSGARGPFPFPGRGVSGPPGPAGPRGANGAPCNDCAKGDAGAPAGSQ	721	
QY	361	GAPGLEMPGRRGAAGLPGPKDRGDAPKAGDSGKQCVRLGTGPIGPPGPAGAPGDK	420	
DB	722	GAPGLEQMPGRRGAAGLPGPKDRGDAGKAGDSGKQCVRLGTGPIGPPGPAGAPGDK	781	
QY	421	GSBSGSPGAGPTGARGAPGDRGEPGPPGPAGTAGPPGADGEPCKAGEPDDAKGSDAGPP	480	
DB	782	GSBSGSPGAGPTGARGAPGDRGEPGPPGPAGTAGPPGADGQPCAKGEPDAGAKGSDAGPP	841	
QY	481	GPAGPAGPPGPIGVGAPGAKARGSGAPGATGFPCAAGRVGPPGCPGADGPPGPPGA	540	
DB	842	GPAGPAGPPGPIGVNGAPGAKARGSGAPGATGFPQAGRVGPPGCGNAGPPGPPGPA	901	
QY	541	GKEG 544		
DB	902	GKEG 905		

RESULT 14
US-10-301-822-28
; Sequence 28, Application US/10301822
; Publication No. US20030148410A1

Fri Sep 24 11:45:39 2004

us-10-658-989a-4.rapb

Page 10

Search completed: September 24, 2004, 11:13:25
Job time : 69.1873 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: September 24, 2004, 11:06:56 ; Search time 20.9509 Seconds
(without alignments)
2497.663 Million cell updates/sec

Title: US-10-658-989A-4

Perfect score: 3070
Sequence: 1 GSGPGEVRCGPPGPPAGA.....PGSPGAGPPGPPGPPAGKEG 544

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283366 seqs, 96191526 residues

Total number of hits satisfying chosen parameters: 283366

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : PIR_78:*
1: PIR1.*
2: PIR2.*
3: PIR3.*
4: PIR4.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2966	96.6	1464	1 CGHUI5	collagen alpha 1(I)
2	2857	93.1	1453	2 S21626	collagen alpha 1(I)
3	2775	90.4	1042	1 CGCHUS	collagen alpha 1(I)
4	2220	72.3	1418	2 T45457	collagen alpha 1(I)
5	2202	71.7	1487	1 CGHJEC	collagen alpha 1(I)
6	2189	71.3	1419	2 A41182	collagen alpha 1(I)
7	2189	71.3	1487	2 B41182	collagen alpha 1(I)
8	2154	70.2	671	1 CGRTIS	collagen alpha 1(I)
9	2147	69.9	1486	1 B40333	collagen alpha 1(I)
10	2131	69.4	1492	2 A40333	collagen alpha 1(I)
11	2076.5	67.6	1779	1 CGBOIS	collagen alpha 1(I)
12	1921.5	62.6	1496	1 CGHJ2V	collagen alpha 2(V)
13	1915	62.4	1373	1 A43231	collagen alpha 2(V)
14	1910	62.2	1497	2 T49607	procollagen type V
15	1908	62.1	1464	2 S59856	collagen alpha 1(I)
16	1905.5	62.1	1049	1 CGBO7S	collagen alpha 1(I)
17	1901	61.9	1466	1 CGHJ7L	collagen alpha 1(I)
18	1886	61.4	1366	1 CGHJ2S	collagen alpha 1(I)
19	1881	61.3	886	2 T50694	collagen alpha 1(I)
20	1550.5	50.5	1414	1 S23809	collagen alpha 2(I)
21	1457.5	47.5	1838	1 CGHUIV	collagen alpha 1(V)
22	1444	47.0	1806	1 CGHUIE	collagen alpha 1(X)
23	1432.5	46.7	1843	2 S18803	collagen alpha 1(V)
24	1418.5	46.2	1027	2 S28774	collagen alpha cha
25	1413	46.0	964	1 CGCHUS	collagen alpha 2(I)
26	1411.5	46.0	1024	2 S18251	collagen alpha 1(X)
27	1400	45.6	888	2 S28791	collagen alpha 1(X)
28	1399.5	45.6	1546	1 CGHJ2E	collagen alpha 2(X)
29	1376	44.8	1691	1 S22917	collagen alpha 5(I)

RESULT 1

CGHUI5

collagen alpha 1(I) chain precursor - human
N;Alternate names: procollagen alpha 1(I) chain

C;Species: Homo sapiens (man)

C;Date: 12-Aug-1981 #sequence_revision 04-Oct-1996 #text_change 31-Dec-2000
C;Accession: I60114; S01143; A93335; I55254; A39943; I55237; A35233; S09400; B90567; S11269; A29439; I53466; A02852; I37247

R;D'Alessio, M.; Bernard, M.; Pretorius, P.J.; de Wet, W.; Ramirez, F.; Pretorius, P.J.
Gene 67, 105-115, 1988

A;Title: Complete nucleotide sequence of the region encompassing the first twenty-five e.
A;Reference number: I60114; MUID:88329734; PMID:2843432

A;Accession: I60114

A;Status: translated from GB/EMBL/DBJ

A;Molecule type: DNA

A;Residues: 1-369, 'L', 371-589 <DAL>

A;Cross-references: GB:M20789; NID:9179593; PIDN:AA859373.1; PID:9179594

R;Tromp, G.; Kuivaniemi, H.; Stacey, A.; Shikata, H.; Baldwin, C.T.; Jaenisch, R.; Prock

Biochem. J. 253, 919-922, 1988

A;Title: Structure of a full-length cDNA clone for the prepro-alpha-1(I) chain of human
A;Reference number: S01143; MUID:89025644; PMID:3178743

A;Accession: S01143

A;Molecule type: mRNA

A;Residues: 1-472 <TRO>

A;Cross-references: EMBL:X07884; NID:G30015; PIDN:CAA30731.1; PID:G30016; GB:M36546; NID

A;Note: submitted to the EMBL/GenBank/DBJ databases by Prockop, D.J., 13-JUN-1988

R;Chu, M.L.; de Wet, W.; Bernard, M.; Ding, J.F.; Morabito, M.; Myers, J.; Williams, C.;

Nature 310, 337-340, 1984

A;Title: Human proalpha1(I) collagen gene structure reveals evolutionary conservation of
A;Reference number: A93335; MUID:84270697; PMID:6462220

A;Accession: A93335

A;Molecule type: DNA

A;Residues: 1-58, 'Q', 60-181 <CHU>

A;Cross-references: EMBL:X00820; NID:G35657; PIDN:CAA23394.1; PID:G35658

R;Rosscow, C.M.S.; Vergeer, W.P.; du Plooy, S.J.; Bernard, M.P.; Ramirez, F.; de Wet, W.

J. Biol. Chem. 262, 15151-15157, 1987

A;Title: DNA sequences in the first intron of the human pro-alpha 1(I) collagen gene enh

A;Reference number: I55254; MUID:88033098; PMID:2822714

A;Accession: I55254

A;Status: translation not shown; translated from GB/EMBL/DBJ

A;Molecule type: DNA

A;Residues: 1-45 <ROS>

A;Cross-references: GB:J02829; NID:G180387; PIDN:AAA51993.1; PID:G180388

R;Bornstein, P.; McKay, J.; Morishima, J.K.; Devarayalu, S.; Gellinas, R.E.

Proc. Natl. Acad. Sci. U.S.A. 84, 8969-8873, 1987

A;Title: Regulatory elements in the first intron contribute to transcriptional control o

A;Reference number: A39943; MUID:88097389; PMID:3480516

A;Accession: A39943

A;Molecule type: DNA

A;Residues: 1-34 <BOR>

A;Cross-references: GB:J03559; NID:G180876; PIDN:AAA52052.1; PID:G553238

R;Chu, M.L.; de Wet, W.; Bernard, M.; Ramirez, F.

J. Biol. Chem. 260, 2315-2320, 1985
A:Title: Fine structural analysis of the human pro-alpha 1 (I) collagen gene. Promoter s
A:Reference number: 155237; MUID:85130970; PMID:2857713
A:Accession: 155237
A:Status: translation not shown; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-34 <CH2>
A:Cross-references: GB:M10627; NID:G180383; PIDN:AAA51992.1; PID:G553226
R:Wirtz, M.K.; Keene, D.R.; Hori, H.; Glanville, R.W.; Steinmann, B.; Rao, V.H.; Hollist
J. Biol. Chem. 265, 6312-6317, 1990
A:Title: In vivo and in vitro noncovalent association of excised alpha 1(I) amino-termina
tome, type VII.
A:Reference number: A35233; MUID:90202908; PMID:2318955
A:Accession: A35233
A:Molecule type: protein
A:Residues: 33-52 <WTR>
A:Note: this propeptide fragment remained non-covalently bound to a defective, uncleaved
E:Wail, D.; d'Alessio, M.; Ramirez, F.; de Wet, W.; Cole, W.G.; Chan, D.; Bateman, J.F.
EMBO J. 8, 1705-1710, 1989
A:Title: A base substitution in the exon of a collagen gene causes alternative splicing
A:Reference number: S09400; MUID:89356643; PMID:2767050
A:Accession: S09400
A:Molecule type: mRNA
A:Residues: 156-183 <WEI>
R:Click, E.M.; Bornstein, P.
Biochemistry 9, 4699-4706, 1970
A:Title: Isolation and characterization of the cyanogen bromide peptides from the alpha
A:Reference number: A90567; MUID:71038625; PMID:5529814
A:Contents: CNBr0-1, CNBr2, CNBr4, CNBr5
A:Accession: B90567
A:Molecule type: protein
A:Residues: 182-198, 'Z', 200-201, 'Z', 203-206, 'Z', 208-209, 'Z', 211-228, 'B', 230, 'BB', 233, 'Z'
A:Experimental source: skin
A:Note: evidence for 170-allysine
R:Baetge, B.; Norbom, H.; Diebold, J.; Lehmann, H.; Bodo, M.; Deutzmann, R.; Mueller, H.
Eur. J. Biochem. 192, 153-159, 1990
A:Title: A critical crosslink region in human-bone-derived collagen type I. Specific cle
A:Reference number: S11372; MUID:90382436; PMID:2169412
A:Accession: S11372
A:Molecule type: protein
A:Residues: 175-187; 274-287, 'P', 289 <BAE>
A:Note: sequence of collagen alpha 1(I) isolated from bone after pepsin digestion
R:Deak, S.B.; Scholz, P.M.; Amenta, P.S.; Constantinou, C.D.; Levi-Minzi, S.A.; Gonzalez
J. Biol. Chem. 266, 21827-21832, 1991
A:Title: The substitution of arginine for glycine 85 of the alpha 1(I) procollagen chain
operative melting of intact type I collagen.
A:Reference number: 155342; MUID:92042092; PMID:1718984
A:Accession: 155342
A:Status: translated from GB/EMBL/DBJ
A:Molecule type: mRNA
A:Residues: 258-268; 1347-1357 <DEA>
A:Cross-references: GB:S67495; NID:9239007; PIDN:AAB20350.1; PID:G239008
A:Note: sequences from the 5' and 3' ends only are shown; mutant sequence 263-Arg report
R:Morgan, P.H.; Jacobs, H.G.; Segrest, J.P.; Cunningham, L.W.
J. Biol. Chem. 245, 5042-5048, 1970
A:Title: Comparative study of glycopeptides derived from selected vertebrate collagens.
A:Reference number: A92069; MUID:71001508; PMID:4319110
A:Accession: A92069
A:Molecule type: protein
A:Residues: 263-268 <MOP>
A:Experimental source: skin
A:Note: attachment of 2-O-alpha-D-glucosyl-O-beta-D-galactose to 5-hydroxylysine
R:Labhard, M.E.; Hollister, D.W.
Matrix 10, 124-130, 1990
A:Title: Segmental amplification of the entire helical and telopeptide regions of the cD
A:Reference number: S15989; MUID:90326017; PMID:2374517
A:Accession: S15989
A:Molecule type: mRNA
A:Residues: 281-302; 402-420; 823-843; 925-944; 1026-1045; 1143-1162 <IAB>
R:Wirtz, M.K.; Rao, V.H.; Glanville, R.W.; Labhard, M.E.; Pretorius, P.J.; de Vries, W.N.
Connect. Tissue Res. 29, 1-11, 1993
A:Title: A cysteine for glycine substitution at position 175 in an alpha 1 (I) chain of
A:Reference number: 152905; MUID:93339042; PMID:8339541
A:Accession: 152905

A:Accession: 152905
A:Status: translated from GB/EMBL/DBJ
A:Molecule type: mRNA
A:Residues: 342-352, 'C', 354-359 <W12>
A:Cross-references: GB:S64717; NID:G408195; PIDN:AAB27677.1; PID:G408196
A:Note: mutant sequence from patient with osteogenesis imperfecta
R:Bernard, M.P.; Chu, M.L.; Myers, J.C.; Ramirez, F.; Eikenberry, E.F.; Prockop, D.J.
Biochemistry 22, 5213-5223, 1983
A:Title: Nucleotide sequences of complementary deoxyribonucleic acids for the proalpha 1
A:Reference number: A90476; MUID:84080385; PMID:6689127
A:Accession: A90476
A:Molecule type: mRNA
A:Residues: 425-1250, 'X', 1252-1328, 'S', 1330-1390, 'X', 1392-1464 <BER>
A:Cross-references: GB:K01228; NID:G180391; PIDN:AAA51995.1; PID:G180392
A:Note: sequence partially completed for missing nucleotides by A29439
R:Chu, M.L.; Gargiulo, V.; Williams, C.J.; Ramirez, F.
J. Biol. Chem. 260, 691-694, 1985
A:Title: Multixon deletion in an osteogenesis imperfecta variant with increased type III
A:Reference number: A22161; MUID:85104934; PMID:2981843
A:Accession: A22161
A:Molecule type: DNA
A:Residues: 472-594, 'R', 596-607 <CH3>
A:Cross-references: GB:K03178; GB:K03179; NID:G179612; NID:G179613; PIDN:AAA51847.1; PID:
A:Note: the authors translated the codon CGT for residue 595 as Pro
R:Wallis, G.A.; Starman, B.J.; Zinn, A.B.; Byers, P.H.
Am. J. Hum. Genet. 46, 1034-1040, 1990
A:Title: Variable expression of osteogenesis imperfecta in a nuclear family is explained
A:Reference number: A35336; MUID:90252792; PMID:2339700
A:Accession: A35336
A:Molecule type: mRNA
A:Residues: 710-720, 'E', 722-737, 'E', 739-745 <WAL>
A:Note: the authors translated the codons CAG for 721 and CGT for 738 as Glu
R:Forlino, A.; Zolezzi, F.; Valli, M.; Pignatti, P.F.; Cetta, G.; Brunelli, P.C.; Mottes,
Hum. Mol. Genet. 3, 2201-2206, 1994
A:Title: Severe (type III) osteogenesis imperfecta due to glycine substitutions in the ce
A:Reference number: 154365; MUID:95187161; PMID:7881420
A:Accession: 154365
A:Status: translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 746-766, 'S', 768-781 <FOR>
A:Cross-references: GB:L47667; NID:G1009093; PIDN:AAB59576.1; PID:G1009094
R:Chessler, S.D.; Wallis, G.A.; Byers, P.H.
J. Biol. Chem. 268, 18218-18225, 1993
A:Title: Mutations in the carboxyl-terminal propeptide of the pro alpha 1(I) chain of tyI
A:Reference number: A47426; MUID:93352646; PMID:8349697
A:Accession: A47426
A:Molecule type: mRNA
A:Residues: 1179-1276, 'H', 1278-1336, 1339-1387, 'R', 1389-1464 <CHE>
A:Cross-references: GB:S64596; NID:G407589; PIDN:AAB27856.1; PID:G407590
A:Note: sequence extracted from NCBI backbone (NCBIN:136444, NCBI:136445)
A:Note: does not represent an experimentally determined sequence but three different muta
A:Accession: B47426
A:Molecule type: mRNA
A:Residues: 1179-1464 <CH4>
A:Experimental source: normal dermal fibroblast culture
A:Accession: C47426
A:Molecule type: mRNA
A:Residues: 1179-1276, 'H', 1278-1464 <CH5>
A:Experimental source: fetal cell 86-237
A:Accession: D47426
A:Molecule type: mRNA
A:Residues: 1179-1336, 1339-1464 <CH6>
A:Experimental source: fetal cell 86-146
A:Accession: E47426
A:Molecule type: mRNA
A:Residues: 1179-1387, 'R', 1389-1464 <CH7>
A:Experimental source: fetal cell 88-251
R:Chom, D.H.; Apone, S.; Eyre, D.R.; Starman, B.J.; Andreassen, P.; Charbonneau, H.; Nicl
J. Biol. Chem. 263, 14605-14607, 1988
A:Title: Substitution of Cysteine for Glycine within the Carboxyl-terminal Telopeptide of
A:Reference number: 155269; MUID:89008319; PMID:3170557
A:Accession: 155269

QY 301 GVPGLGAPGPSGARGEPFGFGRGVGGPPGPAGPGADGAPGDGAKGADAGAPGAPSE 360
| | | | |
Db 615 GVPCEAGAPGLVPRGERGFPEGRSGPQAQLQGARGLPGTPGTDTGPKGASGAPGPPGAQ 674

QY 361 GAPGLEMPGREGAAGLPKPDGRDAGPKGADGSPKGVRGLTPIGPPGAPGAPGDK 420
| | | | |
Db 675 GPPGLQGMPEGREGAAGIAPRGDRGDCVGEKGEPAKDGGRGLTGIFPGFPAGANGEK 734

QY 421 GESPSGPAGPTGARGAPDGRGEPGPPCPAGFAGPCGADGEPGAKGEPGDAGAKGDAGPP 480
| | | | |
Db 735 GEVGPFGPATGATGARGAPGERGETGPPCFAGFAGPGADGPGAKGFCGAGQKGDAGAP 794

QY 481 GPACGAPGPPGPIGDVGAPGAKGARGSGAGPPGATGPFCAARVGPFPSPGSDAGPPGPPGPA 540
| | | | |
Db 795 GFQPSGAPGPPQGTGTGTPKGASGAQGPAGTCFPGAARVGPFGSNGNPNPFPPPGPS 854

QY 541 KKEG 544
| | | | |
Db 855 GKDG 858

RESULT 5
CGHU6C
collagen alpha 1(II) chain precursor [validated] - human
N:Alternate names: procollagen alpha 1(II) chain
C:Contains: chondrocalcin; Collagen alpha 1(II) chain precursor splice form I; collagen
C:Species: Homo sapiens (man)
C:Date: 28-May-1986 #sequence revision 01-Sep-1995 #text change 08-Dec-2000
C:Accession: A38513; S06715; S24270; A24828; S06496; A35428; A30147; A33116; S64674; S637250; I37251; I37252; I37253; I37254; I55338; I59535; I61910
R:Ryan, M.C.; Sieraski, M.; Sandell, L.J.
Genomics 8, 41-48, 1990
A>Title: The human type II procollagen gene: identification of an additional protein-cod
A:Reference number: A38513; MUID:91184811; PMID:2081599
A:Accession: A38513
A:Molecule type: DNA
A:Residues: 1-103 <RY>
A:CROSS-references: GB:M60289; NID:g180883; PIDN:AAA73873.1; PID:g180884
R:Su, M.W.; Lee, B.; Ramirez, F.; Machado, M.; Horton, W.
Nucleic Acids Res. 17, 9473, 1989
A>Title: Nucleotide sequence of the full length cDNA encoding for human type II procolla
A:Reference number: S06715; MUID:90067946; PMID:2587267
A:Accession: S06715
A:Molecule type: mRNA
A:Residues: 1-28, 'R', 99-1487 <SU2>
A:CROSS-references: EMBL:X16468; NID:g29515; PIDN:CAA34488.1; FID:g29516
A>Note: alternative splice form 1
R:Vikkula, M.; Metsäsaaranta, M.; Syvaenen, A.C.; Ala-Kokko, L.; Vuorio, E.; Peltonen, L.
Biochem. J. 285, 287-294, 1992
A>Title: Structural analysis of the regulatory elements of the type-II procollagen gene.
A:Reference number: S24270; MUID:92344585; PMID:1637314
A:Accession: S24270
A>Status: translation not shown
A:Molecule type: DNA
A:Residues: 1-28 <VIK>
A:CROSS-references: EMBL:X58709; GB:S40537; NID:g35659
A>Note: this translation is not annotated in GenBank entry HSPROCOE1, release 111.0
R:Nunez, A.M.; Kohno, K.; Martin, G.R.; Yamada, Y.
Gene 44, 11-16, 1986
A>Title: Promoter region of the human pro-alpha-1-(II)-collagen gene.
A:Reference number: A24828; MUID:87031574; PMID:3021582
A:Accession: A24828
A:Molecule type: DNA
A:Residues: 1-8, 'T', 10-28 <NUN>
A:CROSS-references: GB:M25698; NID:g180872; PIDN:AAAS2051.1; PID:g553237
R:Baldwin, C.T.; Reginaldo, A.M.; Smith, C.; Jimenez, S.A.; Prockop, D.J.
Biochem. J. 262, 521-528, 1989
A>Title: Structure of cDNA clones coding for human type II procollagen. The alpha-1(II)
A:Reference number: S06496; MUID:90026318; PMID:2803268
A:Accession: S06496
A:Molecule type: mRNA
A:Residues: 7-28, 'R', 99-157, 'P', 159-440, 'G', 442-456, 'E', 458-640, 'A', 642-831, 'PA', 834, 'F'

A:Cross-references: EMBL:X16711; NID:g30040; PIDN:CAA34683.1; PID:g30041
A>Note: alternative splice form 1
R:Ryan, M.C.; Sandell, L.J.
J. Biol. Chem. 265, 10334-10339, 1990
A>Title: Differential expression of a cysteine-rich domain in the amino-terminal propept
A:Reference number: A35428; MUID:90285153; PMID:2355003
A:Accession: A35428
A>Status: not compared with conceptual translation
A:Molecule type: mRNA
A:Residues: 27-81, 'L', 83-103 <RYA2>
A>Note: alternative splice form 2; splicing appears to be under developmental regulation
R:Su, M.W.; Benson-Chanda, V.; Vissing, H.; Ramirez, F.
Genomics 4, 438-441, 1989
A>Title: Organization of the exons coding for Pro alpha-1(II) collagen N-propeptide conf
A:Reference number: A30147; MUID:89233138; PMID:2714801
A:Accession: A30147
A:Molecule type: DNA
A:Residues: 104-157, 'P', 159-236 <SUM>
A:Cross-references: GB:J03065; GB:M23660; GB:M25655; GB:M25656; GB:M25730; GB:M32168; GB:M32169; GB:M32170; GB:M32171; GB:M32172; GB:M32173; GB:M32174; GB:M32175; GB:M32176; GB:M32177; GB:M32178; GB:M32179; GB:M32180; GB:M32181; GB:M32182; GB:M32183; GB:M32184; GB:M32185; GB:M32186; GB:M32187; GB:M32188; GB:M32189; GB:M32190; GB:M32191; GB:M32192; GB:M32193; GB:M32194; GB:M32195; GB:M32196; GB:M32197; GB:M32198; GB:M32199; GB:M32200; GB:M32201; GB:M32202; GB:M32203; GB:M32204; GB:M32205; GB:M32206; GB:M32207; GB:M32208; GB:M32209; GB:M32210; GB:M32211; GB:M32212; GB:M32213; GB:M32214; GB:M32215; GB:M32216; GB:M32217; GB:M32218; GB:M32219; GB:M32220; GB:M32221; GB:M32222; GB:M32223; GB:M32224; GB:M32225; GB:M32226; GB:M32227; GB:M32228; GB:M32229; GB:M32230; GB:M32231; GB:M32232; GB:M32233; GB:M32234; GB:M32235; GB:M32236; GB:M32237; GB:M32238; GB:M32239; GB:M32240; GB:M32241; GB:M32242; GB:M32243; GB:M32244; GB:M32245; GB:M32246; GB:M32247; GB:M32248; GB:M32249; GB:M32250; GB:M32251; GB:M32252; GB:M32253; GB:M32254; GB:M32255; GB:M32256; GB:M32257; GB:M32258; GB:M32259; GB:M32260; GB:M32261; GB:M32262; GB:M32263; GB:M32264; GB:M32265; GB:M32266; GB:M32267; GB:M32268; GB:M32269; GB:M32270; GB:M32271; GB:M32272; GB:M32273; GB:M32274; GB:M32275; GB:M32276; GB:M32277; GB:M32278; GB:M32279; GB:M32280; GB:M32281; GB:M32282; GB:M32283; GB:M32284; GB:M32285; GB:M32286; GB:M32287; GB:M32288; GB:M32289; GB:M32290; GB:M32291; GB:M32292; GB:M32293; GB:M32294; GB:M32295; GB:M32296; GB:M32297; GB:M32298; GB:M32299; GB:M32300; GB:M32301; GB:M32302; GB:M32303; GB:M32304; GB:M32305; GB:M32306; GB:M32307; GB:M32308; GB:M32309; GB:M32310; GB:M32311; GB:M32312; GB:M32313; GB:M32314; GB:M32315; GB:M32316; GB:M32317; GB:M32318; GB:M32319; GB:M32320; GB:M32321; GB:M32322; GB:M32323; GB:M32324; GB:M32325; GB:M32326; GB:M32327; GB:M32328; GB:M32329; GB:M32330; GB:M32331; GB:M32332; GB:M32333; GB:M32334; GB:M32335; GB:M32336; GB:M32337; GB:M32338; GB:M32339; GB:M32340; GB:M32341; GB:M32342; GB:M32343; GB:M32344; GB:M32345; GB:M32346; GB:M32347; GB:M32348; GB:M32349; GB:M32350; GB:M32351; GB:M32352; GB:M32353; GB:M32354; GB:M32355; GB:M32356; GB:M32357; GB:M32358; GB:M32359; GB:M32360; GB:M32361; GB:M32362; GB:M32363; GB:M32364; GB:M32365; GB:M32366; GB:M32367; GB:M32368; GB:M32369; GB:M32370; GB:M32371; GB:M32372; GB:M32373; GB:M32374; GB:M32375; GB:M32376; GB:M32377; GB:M32378; GB:M32379; GB:M32380; GB:M32381; GB:M32382; GB:M32383; GB:M32384; GB:M32385; GB:M32386; GB:M32387; GB:M32388; GB:M32389; GB:M32390; GB:M32391; GB:M32392; GB:M32393; GB:M32394; GB:M32395; GB:M32396; GB:M32397; GB:M32398; GB:M32399; GB:M32400; GB:M32401; GB:M32402; GB:M32403; GB:M32404; GB:M32405; GB:M32406; GB:M32407; GB:M32408; GB:M32409; GB:M32410; GB:M32411; GB:M32412; GB:M32413; GB:M32414; GB:M32415; GB:M32416; GB:M32417; GB:M32418; GB:M32419; GB:M32420; GB:M32421; GB:M32422; GB:M32423; GB:M32424; GB:M32425; GB:M32426; GB:M32427; GB:M32428; GB:M32429; GB:M32430; GB:M32431; GB:M32432; GB:M32433; GB:M32434; GB:M32435; GB:M32436; GB:M32437; GB:M32438; GB:M32439; GB:M32440; GB:M32441; GB:M32442; GB:M32443; GB:M32444; GB:M32445; GB:M32446; GB:M32447; GB:M32448; GB:M32449; GB:M32450; GB:M32451; GB:M32452; GB:M32453; GB:M32454; GB:M32455; GB:M32456; GB:M32457; GB:M32458; GB:M32459; GB:M32460; GB:M32461; GB:M32462; GB:M32463; GB:M32464; GB:M32465; GB:M32466; GB:M32467; GB:M32468; GB:M32469; GB:M32470; GB:M32471; GB:M32472; GB:M32473; GB:M32474; GB:M32475; GB:M32476; GB:M32477; GB:M32478; GB:M32479; GB:M32480; GB:M32481; GB:M32482; GB:M32483; GB:M32484; GB:M32485; GB:M32486; GB:M32487; GB:M32488; GB:M32489; GB:M32490; GB:M32491; GB:M32492; GB:M32493; GB:M32494; GB:M32495; GB:M32496; GB:M32497; GB:M32498; GB:M32499; GB:M32500; GB:M32501; GB:M32502; GB:M32503; GB:M32504; GB:M32505; GB:M32506; GB:M32507; GB:M32508; GB:M32509; GB:M32510; GB:M32511; GB:M32512; GB:M32513; GB:M32514; GB:M32515; GB:M32516; GB:M32517; GB:M32518; GB:M32519; GB:M32520; GB:M32521; GB:M32522; GB:M32523; GB:M32524; GB:M32525; GB:M32526; GB:M32527; GB:M32528; GB:M32529; GB:M32530; GB:M32531; GB:M32532; GB:M32533; GB:M32534; GB:M32535; GB:M32536; GB:M32537; GB:M32538; GB:M32539; GB:M32540; GB:M32541; GB:M32542; GB:M32543; GB:M32544; GB:M32545; GB:M32546; GB:M32547; GB:M32548; GB:M32549; GB:M32550; GB:M32551; GB:M32552; GB:M32553; GB:M32554; GB:M32555; GB:M32556; GB:M32557; GB:M32558; GB:M32559; GB:M32560; GB:M32561; GB:M32562; GB:M32563; GB:M32564; GB:M32565; GB:M32566; GB:M32567; GB:M32568; GB:M32569; GB:M32570; GB:M32571; GB:M32572; GB:M32573; GB:M32574; GB:M32575; GB:M32576; GB:M32577; GB:M32578; GB:M32579; GB:M32580; GB:M32581; GB:M32582; GB:M32583; GB:M32584; GB:M32585; GB:M32586; GB:M32587; GB:M32588; GB:M32589; GB:M32590; GB:M32591; GB:M32592; GB:M32593; GB:M32594; GB:M32595; GB:M32596; GB:M32597; GB:M32598; GB:M32599; GB:M32600; GB:M32601; GB:M32602; GB:M32603; GB:M32604; GB:M32605; GB:M32606; GB:M32607; GB:M32608; GB:M32609; GB:M32610; GB:M32611; GB:M32612; GB:M32613; GB:M32614; GB:M32615; GB:M32616; GB:M32617; GB:M32618; GB:M32619; GB:M32620; GB:M32621; GB:M32622; GB:M32623;

A>Note: the composition of peptides comprising residues 1-9 and 1-19 confirms the sequence
A>Note: this region (residues 651-671 above) probably corresponds to positions 1032-1053 (C)
C:Comment: Prolines and lysines at the third position of the tripeptide repeating unit (C)
ed and subsequently O-glycosylated.
C:Comment: The order of the nine CNBr peptides in the alpha 1(I) chain of rat skin collag
C:Superfamily: collagen alpha 1(I) chain; fibrillar collagen carboxyl-terminal homology;
C:Keywords: blocked amino end; coiled coil; extracellular matrix; glycoprotein; hydroxyl
F/I/Modified site: blocked amino end (Glx) probably pyrrolidone carboxylic acid #status
F/I/Modified site: allysine (Lys) #status experimental
F/I/Modified site: 103,424,547/Binding site: carbohydrate (Lys) (covalent) #status experimental
F/I/Modified site: 5-hydroxylysine (Lys) #status experimental
F/I/Modified site: 5-hydroxylysine (Lys) (partial) #status experimental

Query Match 70.2%; Score 2154; DB 1; Length 671;
Best Local Similarity 77.3%; Pred. No. 5.9e-98;
Matches 381; Conservative 28; Mismatches 36; Indels 48; Gaps 2;

QY	1	GSEGEVGRGEPFPAGAPADPGADGPAGKAGDAGPIAGAFGGPARGPSPE	60
DB	200	GSEGQGVGRGEPFPAGAGPAGNPDQQPGAKGANGAGTACPFPAARGPSGQ	259
QY	61	GFGPPPKGSGEFGAPSGSKDGTAKGEGPVVGPPGAGEGKPGARGETPTGLP	120
DB	260	GPSGAPGPKNSGEPAPCNKDTGAKGEPAGVQGPPGABEKGARGGEPFSLP	319
QY	121	GPPEGRRGSPRGFPDAGVAGPKPAGERGSPGAPGKSGPEAGRGEAGLPKAKLT	180
DB	320	GPPEGRRGSPRGFPDAGVAGPKPAGERGSPGAPGKSGPEAGRGEAGLPKAKLT	379
QY	181	GPSGSPGDGKTGPPGAPAGEDRGPFPDPGARGEAGVMGPPGKAAGEPGKAGERVP	240
DB	380	GPSGSPGDGKTGPPGAPGEDBGRPGAPPPCARQAGVMGPPGKTAGEPGKAGERVP	439
QY	241	GPAGVAGXKDGENAGFPPGPAGPAGEREEGPAGSPGFELPGAPGPCEAGKPGEE	300
DB	440	GPAGVAGXKDGENAGFPPGPAGPAGEREEGPAGSPGFELPGAPGPCEAGKPGZZ	499
QY	301	GVPGDLGAPGSGARGEFPFGERGVEGPPGAPGADGAPDDGAKGADAGCAPGSE	360
DB	500	GVPGDLGAPGSGARGERGFGERGVQPPGPAGRNNGAPBBGAKGDTGAEAFGSQ	559
QY	361	GAPGLEHWPERGAAGLPPGKDXRDGAPKADGSPKDGVRGLTGTPGPPGAPAGDK	420
DB	560	GAPGLZGMSGLZ-----GPPSAGSPKBLGLPLGPIGPPGPRGTGBAGPS	631
QY	421	GESGSPGAGTARGAPDRGEPFPAGFAGPPGADGEPGKAGEPGDAGKDGAPP	480
DB	584	GPSNASGPAGPR-----GPSASGSPKBLGLPLGPIGPPGPRGTGBAGPS	631
QY	481	GPAOPAGPPGPIG	493
DB	632	GPPGPPGPPGPPG	644

RESULT 9
B40333
collagen alpha 1(II) chain precursor - African clawed frog
C:Species: Xenopus laevis (African clawed frog)
C:CDate: 10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change 10-Sep-1999
C:Accession: B40333
R/Su M.W.; Suzuki, H.R.; Bleker, J.J.; Solorsh, M.; Ramirez, F.
J. Cell Biol. 115, 565-575, 1991
A>Title: Expression of two nonallelic type II procollagen genes during Xenopus laevis em
A:Reference number: A40333; MUID:92011898; PMID:1918153
A:Accession: B40333
A>Status: preliminary
A:Molecule type: mRNA
A:Residues: 1-1486 <SUA>
A:Cross-references: GB:M63595
C:Superfamily: collagen alpha 1(I) chain; fibrillar collagen carboxyl-terminal homology;
C:Keywords: coiled coil; extracellular matrix; glycoprotein; trimer; triple helix
F:I/37-96/Domains: von Willebrand factor type C repeat homology <WVC>

	1	GSGPVGVEGPPGPAGCAAGPADPGADGDFGAKGDAGPGLTAGAPFFCARGPSQPE	60
QY	:	:::::	:
Dd	:	:::::	:
	389	GPGAQGPRGESGTDSGPSASGNFTDGIPLGAKSGSASGITAGA.GFFGPCRPPCPGF	448
QY	:	:::::	:
Dd	:	:::::	:
	61	GPGGPPPKDSDGEPAPOSKODTAKGBPVVGVGGPPAGABEKGFKAGRGEPFGLP	120
QY	:	:::::	:
Dd	:	:::::	:
	449	GATGPLPKGETGDPIAGFKGEHGPKEIGISAGPQQAGFPAGEEGEKGAAGEPGAAGPL	508
Dd	:	:::::	:
	121	GPGERGGCSRGFPFGADGVAGPKGPAGBGRSPGAPGKGSCEAGRTPGEAGLPGAKGLT	180
QY	:	:::::	:
Dd	:	:::::	:
	509	GPGERGAPNRRGFPQQDLAIPKGA.PBERGVPLGGPKGXNGD.FRTGEPCLFGARGLT	568
Dd	:	:::::	:
	181	GSPGSPDPCKTGPPGPAGEDCRGPPGPPGARGAGVNVGFFGPKGAAGECGKAGERV	240
QY	:	:::::	:
Dd	:	:::::	:
	569	GRPDCAGPQKVGPSAAGEDRGPPGPOGARQQPVGMFFGPKGANGEPGKAGEKGLG	628
Dd	:	:::::	:
	241	GPPGAVGPAKDGEAGAEAPPAGPACBGRGEGEAGSSPFGLPGPAGPPGCEARKPGEE	300
QY	:	:::::	:
Dd	:	:::::	:
	629	GAPRLRLPKDGETCAQCPNGPAGBGRGQGGPPGSGFGLGPDGPPSGEGEGKPDDQ	688
Dd	:	:::::	:
	301	GVPCDLGAPGSPGARGEPGPPGVRGVPFPGAPPGADGAPCDGAKGDAGCAPGSE	360
QY	:	:::::	:
Dd	:	:::::	:
	689	GVPEAGAAGLVNPRGERGFPBGRSGSQQLGSRGLFGT.DGDKPATGFSGPNQAQ	748
Dd	:	:::::	:
	361	GAPLECMPSERGAALPGPKDRGDAGPKGDADSPGDKGVNGLTGPTGPPGPACAGDK	420
QY	:	:::::	:
Dd	:	:::::	:
	749	GPPLQCMPSERGAAGISGSKDRGDTGKBKPEGAFKDGSRGLTGPLGPPGPGPNGEK	808
Dd	:	:::::	:
	421	CESGPSGAPGTGARGAPDRGPPGPPGAPGAPGADGECAGKEPCDCAKGDAGDP	480
QY	:	:::::	:
Dd	:	:::::	:
	809	GESGPSGPPGILVGARGAPDRGENGPPGAPGAPGESPQQAGLKDGQESQKGDAGAP	868
Dd	:	:::::	:
	481	GPAGPAGPPPIGDVGAPGAKGARGSAGPPGATGFCAGAGRVGPPGPGSDAGPPPPGA	540
QY	:	:::::	:
Dd	:	:::::	:
	869	GPQGPSGAPGPPGQTGVNYPKGARGAQGPAGATGFPGAAGRVGTPGPNGNPGPPPPGA	928
Dd	:	:::::	:
	541	KNEG 544	
QY	:		
Dd	:		
	929	KNEG 932	
Dd	:		
	RESULT 11		
	CSB0LS		
	collagen alpha 1(I) chain - bovine (tentative sequence) (fragments)		
	A;Species: Bos primigenius taurus (cattle)		
	A;C.Date: 24-Apr-1984 Sequence Revision 31-Dec-1993 #text_change 31-Mar-2000		
	A;Accession: A91193; A91229; A91387; A91211; A91201; A91200; A43048; AC2853		
	R;Rauberberg, J.; Temp, R.; Furthmayr, H.		
	Eur. J. Biochem. 27, 231-237, 1972		
	A;Title: Structural characterization of N-terminal antigenic determinants in c		
	A;Reference number: A91193; MUID:72255334; PMID:4115172		
	A;Accession: A91193		
	A;Molecule type: protein		
	A;Residues: 1-19 <RAU>		
	A;Experimental source: skin		
	A;Note: The epsilon carbon of Lys-9, by homology with the rat alpha 1(I) chain		
	R;Fietzek, P.P.; Kuehn, K.		
	Eur. J. Biochem. 52, 77-82, 1975		
	A;Title: The covalent structure of collagen: amino-acid sequence of the cyanogly		
	A;Reference number: A91229; MUID:76022320; PMID:1164916		
	A;Accession: A91229		
	A;Molecule type: protein		
	A;Residues: 20-145 <FIE>		
	A;Experimental source: skin		
	A;Note: Lys-103 is hydroxylated and binds glucosylgalactose		
	R;Fietzek, P.P.; Wendt, P.; Kell, I.; Kuehn, K.		
	FEBS Lett. 26, 74-76, 1972		
	A;Title: The covalent structure of collagen: amino acid sequence of alpha1-CB3		
	A;Reference number: A91387; MUID:73049499; PMID:4673951		
	A;Accession: A91387		
	A;Molecule type: protein		
	A;Residues: 146-294 <FIZ>		
	A;Experimental source: skin		

R;Fietzek, P.P.; Rexrodt, F.W.; Hopper, K.E.; Kuehn, K.
 Eur. J. Biochem. 38, 396-400, 1973
 A:Title: The covalent structure of collagen. 2. The amino-acid sequence of alpha1-CB7 fi
 A:Reference number: A91211; MUID:74086118; PMID:4359390
 A:Accession: A91211
 A:Molecule type: protein
 A:Residues: 295-562 <FI3>
 A:Experimental source: skin
 R;Wendt, P.; Mark, K.V.D.; Rexrodt, F.; Kuehn, K.
 Eur. J. Biochem. 30, 169-183, 1972
 A:Title: The covalent structure of collagen. The amino-acid sequence of the 112 residues
 A:Reference number: A91201; MUID:73042276; PMID:4343808
 A:Accession: A91201
 A:Molecule type: protein
 A:Residues: 563-675 <WEN>
 A:Experimental source: skin
 R;Fietzek, P.P.; Rexrodt, F.W.; Wendt, P.; Stark, M.; Kuehn, K.
 Eur. J. Biochem. 30, 163-168, 1972
 A:Title: The covalent structure of collagen. Amino acid sequence of peptide alpha1-CB6-C
 A:Reference number: A91200; MUID:73042275; PMID:4343807
 A:Accession: A91200
 A:Molecule type: protein
 A:Residues: 676-758 <FI4>
 A:Experimental source: skin
 A:Note: Pro-726 is the only 3-hydroxyproline and the only hydroxylated proline in positio
 R;Rauterberg, J.; Fietzek, P.; Rexrodt, F.; Becker, U.; Stark, M.; Kuehn, K.
 FEBS Lett. 21, 75-79, 1972
 A:Title: The amino acid sequence of the carboxyterminal nonhelical cross link region of
 A:Reference number: A43048
 A:Accession: A43048
 A:Molecule type: protein
 A:Residues: 759-779 <RA2>
 A:Experimental source: skin
 C:Comment: Lysines at positions 115, 124, 274, 346, 424, 496, 658, and 670 may be hydrox
 C:Comment: Prolines in the third position of the tripeptide repeating unit (G-X-Y) are h
 C:Comment: The order of the eight CNBr peptides in the alpha 1(I) chain of bovine skin c
 9, 149, 268, and 217 residues.
 C:Comment: The complete chain contains 1052 residues.
 C:Superfamily: collagen alpha 1(I) chain; fibrillar collagen carboxyl-terminal homology;
 C:Keywords: coiled coil; extracellular matrix; glycoprotein; pyroglyutamic acid; trimer;
 P:1/Modified site: pyrrolidone carboxylic acid (Gln) #status experimental
 Query Match 67.6%; Score 2076.5; DB 1; Length 779;
 Best Local Similarity 78.0%; Pred. No. 3.7e-94;
 Matches 382; Conservative 18; Mismatches 51; Indels 39; Gaps 6;
 QY 55 GPSGPEGPGPPGKGSDEGPGAGSGDGTGAKGEGPVGVEGPPGPGAGEGKPGARGEP 114
 Db 17 GFMGPGPRGLFPGP-----PGA-----PGQFQGGP---GEPGEPGASGPM 55
 QY 115 GPTGLPFPGERGPGSRGPPGADGVAGKPGPAGERSGPGAPGKSGPGEAGRPAGLP 174
 Db 56 GPRGPPGPKNGD-----DGEAKGPRGPRGPPGQAR-----GLPGTAGLP 100
 QY 175 GAKLTGSPGSPDGTGTPGAGEDGRPGPPGPGARAGEAGVMPGPPKGAAGEPGXA 234
 Db 101 GMKGHRGFSGLDGAAGDAGPAGPKGPGSGENGAQGMGPRGLPGFPFGKGAAGEPGXA 160
 QY 235 GERGVPPGPAVAGPAGKDGAGAGGPPGAPGAPGAGERGEGPAGSGPFGELGPGAPGGA 294
 Db 161 GERGVPPGPAVAGPAGKDGAGAGGPPGAPGAPGAGERGEGPAGSGPFGELGPGAPGGA 220
 QY 295 GKPGEEGVPDLCAGPAGSGARGPPGPPGPRGVEGPPGAPGADGAPGDGAKGDAGAP 354
 Db 221 GKPGEEGVPDLCAGPAGSGARGPPGPPGPRGVEGPPGAPGADGAPGDGAKGDAGAP 280
 QY 355 GAFSGEAGLEGMPGPRGAGGPGPKGDRGAGKPGAGDSCKDVRGLTPIGPPGPA 414
 Db 281 GAFSGEAGLEGMPGPRGAGGPGPKGDRGAGKPGAGDSCKDVRGLTPIGPPGPA 340
 QY 415 GAPGDKGEGSPGAPGTGARGAGDRGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPG 474
 Db 341 GAPGDKGEGSPGPA---GTRGAPGDRGPPGPPGPPGPPGPPGPPGPPGPPGPPG 397

QY 475 GDAGPPGAPGAPGPPGPIGVGAPGAKGARGAGAGPGATGTFGAAAGRVGPPGPGSDAGPP 534
 Db 398 GDAGPPGAPGAPGPPGPIGVGAPGAKGARGAGAGPGATGTFGAAAGRVGPPGPGSDAGPP 457
 QY 535 GPPGAPGAPGAPGAPGAPGAPGAPGAPGAPGAPGAPGAPGAPGAPGAPGAPGAPG 544
 Db 458 GPPGAPGAPGAPGAPGAPGAPGAPGAPGAPGAPGAPGAPGAPGAPGAPGAPGAPG 467
 RESULT 12
 CGHU2V
 collagen alpha 2(V) chain precursor - human
 C:Species: Homo sapiens (man)
 C:Date: 31-Jul-1989 #sequence revision 28-Jul-1995 #text change 31-Dec-2000
 C:Accession: A31427; A54555; A54643; A25874; I55239; I59025; A25374; A30017
 R;Woodbury, D.; Benson-Chanda, V.; Ramirez, F.
 J. Biol. Chem. 264, 2735-2738, 1989
 A:Title: Amino-terminal propeptide of human pro-alpha2(V) collagen conforms to the struct
 A:Reference number: A31427; MUID:89123368; PMID:2914927
 A:Accession: A31427
 A:Molecule type: mRNA
 A:Residues: 1-463 <WOO>
 A:Cross-references: GB:J04478; NID:gl79697; PIDN:AAA51859.1; PID:gl79698
 A:Experimental source: placenta
 R;Greenspan, D.S.; Lee, S.T.; Lee, B.S.; Hoffman, G.G.
 Gene Expr. 1, 29-39, 1991
 A:Title: Homology between alpha2(V) and alpha1(III) collagen promoters and evidence for r
 A:Reference number: A54555; MUID:92314691; PMID:1820205
 A:Accession: A54555
 A:Molecule type: DNA
 A:Residues: 1-32 <GRE>
 A:Cross-references: GB:M58529; NID:gl80834; PIDN:AAC41699.1; PID:G553235
 R;Moradi-Ameli, M.; Rousseau, J.C.; Kleman, J.P.; Champilaud, M.F.; Boutillon, M.M.; Ber
 Eur. J. Biochem. 231, 987-995, 1994
 A:Title: Diversity in the processing events at the N-terminus of type-V collagen.
 A:Reference number: S43642; MUID:94237164; PMID:8181482
 A:Accession: S43643
 A:Molecule type: protein
 A:Residues: 288-291, 'P', 293-294, 'X', 296-297, 606, 'X', 608-617 <MOR>
 R;Weil, D.; Bernard, M.; Gargano, S.; Ramirez, F.
 Nucleic Acids Res. 15, 181-198, 1987
 A:Title: The pro alpha 2(V) collagen gene is evolutionarily related to the major fibrill
 A:Reference number: A25874; MUID:87146331; PMID:3029669
 A:Accession: A25874
 A:Molecule type: mRNA; DNA
 A:Residues: 398-1496 <WEI>
 A:Cross-references: GB:X04758; NID:G29588; PIDN:CAA28454.1; PID:gl340175
 A:Experimental source: rhabdomyosarcoma cell line
 R;Myers, J.C.; Loidl, H.R.; Stolle, C.A.; Seyer, J.M.
 J. Biol. Chem. 260, 5533-5541, 1985
 A:Title: Partial covalent structure of the human alpha 2 type V collagen chain.
 A:Reference number: I55239; MUID:85182703; PMID:2985598
 A:Accession: I55239
 A:Status: translated from GB/EMBL/DBJ
 A:Molecule type: mRNA
 A:Residues: 1002-1226 <BE2>
 A:Cross-references: GB:M10956; NID:gl80427; PIDN:AAA52007.1; PID:gl80428
 A:Note: part of this sequence were determined by protein sequencing
 R;Emmanuel, B.S.; Cammizaro, L.A.; Seyer, J.M.; Myers, J.C.
 Proc. Natl. Acad. Sci. U.S.A. 82, 3385-3389, 1985
 A:Title: Human alpha 1(III) and alpha 2(V) procollagen genes are located on the long arm
 A:Reference number: I59025; MUID:85216505; PMID:3858826
 A:Accession: I59025
 A:Status: translated from GB/EMBL/DBJ
 A:Molecule type: mRNA
 A:Residues: 1003-1034 <RES>
 A:Cross-references: GB:M11335; NID:gl79693; PIDN:AAA51857.1; PID:gl79694
 A:Note: part of this sequence were determined by protein sequencing
 R;Myers, J.C.; Loidl, H.R.; Seyer, J.M.; Dion, A.S.
 J. Biol. Chem. 260, 11216-11222, 1985
 A:Title: Complete primary structure of the human alpha-2 type V procollagen COOH-terminal
 A:Reference number: A25374; MUID:85289337; PMID:2411731

QY	121	GPVGGGSGSGFFGCDVAGVCKGPAAGRGSPGAGKGSFGAGRGCEAGLPGAKGLT	180
DB	513	GPVGGGAGPKNRGGFFGSDGLPGFGKAQGERGVPVSGSPKGSQCDGFRDGPGLPGARGLT	572
QY	181	GSPGSGPDPDKTGPPCPAGEDGRPGFPDPPGARGAGVYNGFPFGKAAAGEPCKAAGERVVP	240
DB	573	GNPVGQPEGKGLPLGAPGEDGRPGFPDPSIGTKGQPGMTGLFPKPKSGNDPCKPGEAGNP	632
QY	241	GSPGAVGPAKGDEAGAGCPGPGAPAGRGREGPAGSPGPGPLGPGAPGPEAGKPGEE	300
DB	633	GVFGQSGAGFGKQGVGYGFPDPLRGREGGQFPPTGFGCHFGPFGPPGREGKPGDQ	692
QY	301	GVPGDLGAPSGARGEPGFPGERGVGEPGPPAGPPGADGAFDGDGAKGDAGAPAGPGE	360
DB	693	GVFGPGGAVGVLGPRGERGNPGERGEPGITGLFGEKMGAGCHGPDGPKSGPFGSGTPGDT	752
QY	361	GAPGLEGMDEGSAAGLPGKPGDGDAGPKGADGSGPKDGVGLGTGPIGPPGAPAGPDK	420
DB	753	GPPLGCGMPEERGIAGTPGKDRGIGKEKGAEAGTAGNDGAGLPGPLGPPGAGLLGEK	812
QY	421	GSGPSGPGAPGTARGAPGDRGEPGPPGAPGAPGADGEPGAKGEPGDGAKGDAGPP	480
DB	813	GEPPRGLVFPGSRGNPSRGENGPTGAVGAPGQSGDQGVKGEPEPQCKGDAGSP	872
QY	481	GPAGPAGPGPPTGDVGAPCAKAGSAGSPGATGFPFGAAGRVGPPGPGSDAGPPGPGPA	540
DB	873	GPGLAGSPGPHGPNVGPLKGKRGIQGPGATGTPGSGRVPVPPGAPGAPGFLGEP	932
QY	541	KGEG 544	
DB	933	KGEG 936	

RESULT 13
 A43291
 collagen alpha 2(I) chain precursor - mouse
 C:Species: Mus musculus (house mouse)
 C:Date: 10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change 10-Sep-1999
 C:Accession: A43291 A54328
 R:Phillips, C.L.; Morgan, A.L.; Lever, L.W.; Wenstrup, R.J.
 Genomics 13, 1345-1346 1992
 A:Title: Sequence analysis of a full-length cDNA for the murine pro alpha 2(I) collagen
 A:Reference number: A43291; MUID:92372043; PMID:1505972
 A:Accession: A43291
 A:Status: preliminary; not compared with conceptual translation
 A:Molecule type: mRNA
 A:Residues: 1-1373 <PH1>
 A:Cross-references: NID:G50488; PIDN:CAA41205.1; PID:G50489
 A:Note: sequence extracted from NCBI backbone (NCEIP.112027)
 R:Phillips, C.L.; Lever, L.W.; Pinnell, S.R.; Quarles, L.D.; Wenstrup, R.J.
 J. invest. Dermatol. 97, 980-984, 1991
 A:Title: Construction of a full-length murine Proalpha2(I) collagen cDNA by the polymerase
 A:Reference number: A54328; MUID:92084969; PMID:1748823
 A:Accession: A54328
 A:Status: preliminary; not compared with conceptual translation
 A:Molecule type: mRNA
 A:Residues: 1-110 <PH2>
 C:Genetics:
 A:Gene: COL1A2
 C:Superfamily: collagen alpha 2(I) chain; fibrillar collagen carboxyl-terminal homology

Query Match 62.4%; Score 1915; DB 1; Length 1373;
 Best Local Similarity 64.4%; Pred. No. 3.7e-86;
 Matches 349; Conservative 45; Mismatches 150; Indels 0; Gaps 0;

Qy	1	GSGPGVGRGPPG	PAGAGAGAPG	PDGDFGAKGADG	AGCAGAPG	PGHGRSGSPE	60
Db	280	GPAGPAGPGEVGL	GLSGVGGPFG	PNGLTGAKGATGL	PGVAGAPGL	FGPRGTPGFA	339
Qy	61	GPQGPFGKDGSE	PGAPSGKGTG	AKGFPVGVGEP	PGPAGKEG	KPGARGPFGTGLP	120

F;1145-1373/Domain: fibrillar collagen carboxyl-terminal homology <FCC>

```

Db      340 GAAGTATGAGLVLGPPGAGSGKESGNKGPVAGQPPGSGEGKKGSGPGEAGSAGPA 399
QY      121 GPPGRRGGSGRFFGADGAVGPKGAGRGSPGPKGSPGAGRGPGAGRGAGLPCAKGLT 180
Db      400 GPPGLRGSPGSLGADGAGVWGPFGNRSSTGPGIRGPNAGDAGRPCEGLMPPRLP 459
QY      181 GSPGSGPDGKTGPPGAGDGRPPGPPGARGGAVWGPFGKGAERGKAGRGV 240
Db      460 GSPGVGSGKGPVGLPCIDRGPPIGAPRGAGNIGFPKPGSGDPKPGERGH 519
QY      241 GPPGAVGAGKDEAGAGAGPPGAPGAGRGEEGPPAGSPGEGLPAGPPGEACKPGE 300
Db      520 GLAGARGAGPDGNNAGQPPGQVCGKQEGQAGPPGFGQLGPPSGTTGVEVKGER 579
QY      301 GVPGLGARGSGARGGPPGPPGVEGPPGPPGAGPPGADGAGDDGAKGDAGAGPGE 360
Db      580 GLPGEFLGPPGAPRGERTGPGESGAAGSPGPIGSRGSPGAPGPDGKNGEAGAVGAPCSA 639
QY      361 GAPGLEGMPPGGAAGLPGPKGDRDAGPKGADGSPGKDGVRGLTGTGPPGPPAGAPGDK 420
Db      640 GASGPGGLPGERGAGIIPGKGEKGETGLRGDTGNTGFDGARGIPGAVGAPGAGSGDR 699
QY      421 GESGSPGAPGTGARGAGDGRGEPGPPGAGPPGADGPPGAGKGFDDGAKGDAGPP 480
Db      700 GEAGAAGSPGAPRGSPGGERGVEGPPGPPGPPGAGPAGAGQPGAKGKGTGKPGKENGIV 759
QY      481 GPAGPAGPPGPIGDVGAAGKARGSGAGPPGATGPPGAGRVGPPGSGDAGPPGPPGA 540
Db      760 GPTGSGVGAAGSGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGAA 819
QY      541 GKEG 544
Db      820 GKEG 823

RESULT 14
I49607
procollagen type V alpha 2 - mouse
C:Species: Mus musculus (house mouse)
C>Date: 02-Jul-1996 #sequence_revision 02-Jul-1996 #text_change 13-Aug-1999
C:Accession: I49607
R:Andrikopoulos, K.; Suzuki, H.R.; Solursh, M.; Ramirez, F.
Dev. Dyn. 195, 113-120, 1992
A:Title: Localization of pro-alpha 2(V) collagen transcripts in the tissues of the devel
A:Reference number: I49607; MUID:93214071; PMID:1297453
A:Accession: I49607
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: mRNA
A:Cross-references: GB:I02918; NID:G309180; PIDN:AAA37440.1; PID:G309181
C:Genetics:
A:Gene: Col5a-2
C:Superfamily: collagen alpha 1(I) chain; fibrillar collagen carboxyl-terminal homology;
F:39-98/Domain: von Willebrand factor type C repeat homology <VWC>
F:1270-1497/Domain: fibrillar collagen carboxyl-terminal homology <FCC>

Query Match 62.2%; Score 1910; DB 2; Length 1497;
Best Local Similarity 63.8%; Pred. No. 6.9e-86;
Matches 347; Conservative 39; Mismatches 158; Indels 0; Gaps 0;

QY      1 GSEGEVGRGPPGPPGAGAGPAGDGPAGDGPAGKAGDAGPAGPAGPAGPAGPAGPAGP 60
Db      394 GPEGPGQGRGTGPPGAGSGLGAVGTGTGTPKATGATGATGATGATGATGATGATGAT 453
QY      61 GPGGPPGKSGSPGAPGSKGDTGAKGEGPCPVGVEGPPGPPGAGGKPGARGPPGTPGLP 120
Db      454 GSTPQGIINGSGDPPGVFGKGEAGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPP 513
QY      121 GPPGERGGPPGPPGADGVAGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPG 180
Db      514 GPMGERGAPNRPFGSDGLPGKAGAGRGPPGSSGKGGQDPPGPPGPPGPPGPPGPPG 573
QY      181 GSPGSPGPDGKTGPPGAGDGRPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPP 240

```

```

Db      574 GNPVGQPPGKGLPLGAPGEDGRPPGSGIRGQPGSMGVGPPGSGSDGLKPGEGNA 633
QY      241 GPPGAVGPPAGKDEAGAGPPGPPGAGRGEGPAGSPGFEGLPQPPAGPPGAGKPGEE 300
Db      634 GVPQGRGAGKDEGVPSPVPPGGLAGRGAGPPGTGTGFGQLGPPGPPGPPGPPGPPG 693
QY      301 GVPGLGARGSGARGGPPGPPGVEGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPP 360
Db      694 GVPGEFAGVPLGPPRGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPP 753
QY      361 GAPGLEGMPPGGAAGLPGPKGDRDAGPKGADGSPGKDGVRGLTGTGPPGPPAGAPGDK 420
Db      754 GPPGLQGMPPGEGRIAGTGPFGKDRGIGEGKAGGTAGNDGARGLPGPLGPPGAGLLGAP 813
QY      421 GESGSPGAPGTGARGAGDGRGEPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPP 480
Db      814 GPPGPRGLVPPGSRGNPGSRGNGTGAAGVAGPAGQGGQPPGVKGEPPGPPGPPGPP 873
QY      481 GPAGPAGPPGPIGDVGAAGKARGSGAGPPGATGPPGAGRVGPPGSGDAGPPGPPGA 540
Db      874 GPPGLGAGSPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPG 933
QY      541 GKEG 544
Db      934 GKEG 937

RESULT 15
S59856
collagen alpha 1(III) chain precursor - mouse
C:Species: Mus musculus (house mouse)
C>Date: 10-Apr-1996 #sequence_revision 19-Apr-1996 #text_change 13-Aug-1999
C:Accession: S59856; S62120; S16373
R:Toman, P.D.; de Crombrughe, B.
Gene 147, 161-168, 1994
A:Title: The mouse type-III procollagen-encoding gene: genomic cloning and complete DNA s
A:Reference number: S59856; MUID:95011609; PMID:7926795
A:Accession: S59856
A:Molecule type: DNA
A:Residues: 1-1464 <TOM>
A:Cross-references: EMBL:X52046
R:Toman, D.
submitted to the EMBL Data Library, November 1994
A:Reference number: S62120
A:Accession: S62120
A:Molecule type: DNA
A:Residues: 1-866, G' 868-1464 <TOM>
A:Cross-references: EMBL:X52046; NID:G575321; PIDN:CAA36279.1; PID:G575322
R:Metzgerant, M.; Toman, D.; de Crombrughe, B.; Vuorio, E.
Biochim. Biophys. Acta 1089, 241-243, 1991
A:Title: Specific hybridization probes for mouse type I, II, III and IX collagen mRNAs.
A:Reference number: S16176; MUID:91274355; PMID:2054384
A:Accession: S16373
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1442-1464 <MET>
A:Cross-references: EMBL:X57983; NID:G50476; PIDN:CAA11048.1; PID:G50477
C:Genetics:
A:Insertions: 29/1; 95/3; 112/3; 150/3; 175/3; 193/3; 211/3; 229/3; 247/3; 265/3; 283/3; 298/
58/3; 673/3; 706/3; 742/3; 760/3; 778/3; 796/3; 814/3; 850/3; 866/3; 886/3; 940/3; 976/3;
C:Superfamily: collagen alpha 1(II) chain; fibrillar collagen carboxyl-terminal homology;
C:Keywords: coiled coil; extracellular matrix
F:1-24/Domain: signal sequence #status predicted <SIG>
F:25-154/Domain: propeptide #status predicted <PRO>
F:32-92/Domain: von Willebrand factor type C repeat homology <VWC>
F:155-1464/Product: collagen alpha 1(III) chain #status predicted <MAT>
F:1236-1464/Domain: fibrillar collagen carboxyl-terminal homology <FCC>

Query Match 62.1%; Score 1908; DB 2; Length 1464;
Best Local Similarity 63.3%; Pred. No. 8.5e-86;
Matches 346; Conservative 37; Mismatches 158; Indels 6; Gaps 2;

```

Search completed: September 24, 2004, 11:11:47
Job time : 22.9509 secs

This Page Blank (uspto)

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: September 24, 2004, 11:06:55 ; Search time 13.7264 Seconds
(without alignments)
2063.622 Million cell updates/sec

Title: US-10-658-989A-4
Perfect score: 3070
Sequence: 1 GSEGPGEVGRGPPGPAGA.....PGPSGDAGPRGPPGACKEG 544

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 141681 seqs, 52070155 residues

Total number of hits satisfying chosen parameters: 141681

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SwissProt_42.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	2966	96.6	1464	1	CA11_HUMAN
2	2940	95.8	1460	1	CA11_CANFA
3	2857	93.1	1453	1	CA11_MOUSE
4	2775	90.4	1453	1	CA11_CHICK
5	2202	71.7	1418	1	CA12_HUMAN
6	2192	71.4	1459	1	CA12_MOUSE
7	2151	70.1	671	1	CA11_RAT
8	2076.5	67.6	779	1	CA11_BOVIN
9	1949	63.5	1364	1	CA21_BOVIN
10	1934.5	63.0	1262	1	CA13_CHICK
11	1834	63.0	1372	1	CA21_RAT
12	1921.5	62.6	1496	1	CA25_HUMAN
13	1915	62.4	1372	1	CA21_MOUSE
14	1905.5	62.1	1049	1	CA13_BOVIN
15	1901	61.9	1466	1	CA13_HUMAN
16	1899	61.9	1464	1	CA13_MOUSE
17	1898	61.8	1366	1	CA21_CANFA
18	1880	61.2	1366	1	CA21_HUMAN
19	1834.5	59.8	1355	1	CA21_RANCA
20	1822	59.3	1362	1	CA21_CHICK
21	1789	58.3	1356	1	CA21_ONCMY
22	1496	48.7	747	1	CA12_BOVIN
23	1454.5	47.4	1838	1	CA15_HUMAN
24	1449	47.2	1806	1	CA15_HUMAN
25	1447	47.1	1745	1	CA15_HUMAN
26	1432	46.6	1804	1	CA15_MOUSE
27	1423.5	46.4	1736	1	CA2B_HUMAN
28	1414.5	46.1	1736	1	CA2B_MOUSE
29	1410	45.9	1027	1	CAPP_RIPPA
30	1387	45.2	1685	1	CA5A_HUMAN
31	1375.5	44.8	2944	1	CA17_HUMAN
32	1372	44.7	1763	1	CA24_ASCSU
33	1357.5	44.2	1690	1	CA44_HUMAN
34	1356	44.2	911	1	CA1B_BOVIN
35	1344.5	43.8	1670	1	CA34_HUMAN
36	1339.5	43.6	1603	1	CA1F_HUMAN
37	1320	43.0	1669	1	CA14_HUMAN
38	1310.5	42.7	1758	1	CA24_CAEEL
39	1310	42.7	1669	1	CA14_MOUSE
40	1286.5	41.9	1712	1	CA24_HUMAN
41	1281.5	41.7	1516	1	CA1H_HUMAN
42	1279.5	41.7	754	1	CA54_CANFA
43	1277	41.6	684	1	CA39_HUMAN
44	1276.5	41.6	1143	1	CA1I_HUMAN
45	1276	41.6	921	1	CA19_MOUSE

ALIGNMENTS

RESULT 1
CALL_HUMAN
ID CALL_HUMAN STANDARD; PRT; 1464 AA.
AC P02452; Q14037; Q15176;
DT 21-JUL-1986 (Rel. 01, Created)
DT 01-MAR-1989 (Rel. 10, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Collagen alpha 1(I) chain precursor.
GN COL1A1.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE OF 1-472 FROM N.A.
RX MEDLINE=89025644; PubMed=3178743;
RA Tromp G., Kuivaniemi H., Stacey A., Shikata H., Baldwin C.T.,
RA Jaenisch R., Prockup D.J.;
RT "Structure of a full-length cDNA clone for the prepro alpha 1(I)
RT chain of human type I procollagen.";
RL Biochem. J. 253:919-922(1988).
RN [2]
RP SEQUENCE OF 1-181 FROM N.A.
RX MEDLINE=84270697; PubMed=6462220;
RA Chu M.-L., de Wet W.J., Bernard M.P., Ding J.-F., Morabito M.,
RA Myers J., Williams C., Ramirez F.;
RT "Human pro alpha 1(I) collagen gene structure reveals evolutionary
RT conservation of a pattern of introns and exons.";
RL Nature 310:337-340(1984).
RN [3]
RP SEQUENCE OF 162-301.
RC TISSUE=Skin;
RX MEDLINE=71038625; PubMed=5529814;
RA Click E.M., Bornstein P.;
RT "Isolation and characterization of the cyanogen bromide peptides from
RT the alpha 1 and alpha 2 chains of human skin collagen.";
RL Biochemistry 9:4699-4706(1970).
RN [4]
RP SEQUENCE OF 263-268.
RC TISSUE=Skin;
RX MEDLINE=71001508; PubMed=4319110;
RA Morgan P.H., Jacobs H.G., Segrest J.P., Cunningham L.W.;
RT "A comparative study of glycopeptides derived from selected
RT vertebrate collagens. A possible role of the carbohydrate in fibril
RT formation.";
RL J. Biol. Chem. 245:5042-5048(1970).
RN [5]
RP SEQUENCE OF 425-1464 FROM N.A.
RX MEDLINE=84080385; PubMed=6689127;
RA Bernard M.P., Chu M.-L., Myers J.C., Ramirez F., Eikenberry E.F.,
RA Prockup D.J.;
RT "Nucleotide sequences of complementary deoxyribonucleic acids for the
RT pro alpha 1 chain of human type I procollagen. Statistical evaluation
RT of structures that are conserved during evolution.";
RL Biochemistry 22:5213-5223(1983).
RN [6]

RP SEQUENCE OF 1229-1454 FROM N.A.
RC TISSUE-Bone;
RX MEDLINE=86124208; PubMed=3340531;
RA Maekelae J.K., Raasina M., Virta A., Vuorio E.;
RT "Human pro alpha 1(I) collagen: cDNA sequence for the C-propeptide
RT domain.";
RL Nucleic Acids Res. 16:349-349(1988).
RN [7]
RP SEQUENCE OF 1-34 FROM N.A.
RX MEDLINE=88097389; PubMed=3480516;
RA Bornstein P., McKay J., Morishima J.K., Devarayalu S., Gelinas R.E.;
RT "Regulatory elements in the first intron contribute to
RT transcriptional control of the human alpha 1(I) collagen gene.";
RL Proc. Natl. Acad. Sci. U.S.A. 84:8869-8873(1987).
RN [8]
RP SEQUENCE OF 1-34 FROM N.A.
RX MEDLINE=86130970; PubMed=2857713;
RA Chu M.-L., de Wet W.J., Bernard M.P., Ramirez F.;
RT "Fine structural analysis of the human pro-alpha 1 (I) collagen gene.
RT Promoter structure, Alu repeats, and polymorphic transcripts.";
RL J. Biol. Chem. 260:2315-2320(1985).
RN [9]
RP SEQUENCE OF 1-44 FROM N.A.
RX MEDLINE=88033098; PubMed=2822714;
RA Rosouw C.M.S., Vergeer W.P., du Plooy S.J., Bernard M.P., Ramirez F.,
RA de Wet W.J.;
RT "DNA sequences in the first intron of the human pro-alpha 1(I)
RT collagen gene enhance transcription.";
RL J. Biol. Chem. 262:15151-15157(1987).
RN [10]
RP REVIEW ON VARIANTS.
RX MEDLINE=91184577; PubMed=2010058;
RA Kuvaliemi H., Tromp G., Prockop D.J.;
RT "Mutations in collagen genes: causes of rare and some common diseases
RT in humans.";
RL FASEB J. 5:2052-2060(1991).
RN [11]
RP REVIEW ON VARIANTS.
RX MEDLINE=97255959; PubMed=9101290;
RA Kuvaliemi H., Tromp G., Prockop D.J.;
RT "Mutations in fibrillar collagens (types I, II, III, and XI), fibril-
RT associated collagen (type IX), and network-forming collagen (type X)
RT cause a spectrum of diseases of bone, cartilage, and blood vessels.";
RL Hum. Mutat. 9:300-315(1997).
RN [12]
RP REVIEW ON OI VARIANTS.
RX MEDLINE=91374476; PubMed=1895312;
RA Byers P.H., Wallis G.A., Willing M.C.;
RT "Osteogenesis imperfecta: translation of mutation to phenotype.";
RL J. Med. Genet. 28:433-442(1991).
RN [13]
RP REVIEW ON OI VARIANTS.
RX MEDLINE=97169389; PubMed=9016532;
RA Daigleish R.;
RT "The human type I collagen mutation database.";
RL Nucleic Acids Res. 25:181-187(1997).
RN [14]
RP VARIANT OI-II CYS-1166.
RX MEDLINE=86287390; PubMed=3016737;
RA Conn D.H., Byers P.H., Steinmann B., Gelinas R.E.;
RT "Lethal osteogenesis imperfecta resulting from a single nucleotide
RT change in one human pro alpha 1(I) collagen allele.";
RL Proc. Natl. Acad. Sci. U.S.A. 83:6045-6047(1986).
RN [15]
RP VARIANT OI-II CYS-569.
RX MEDLINE=87222295; PubMed=3108247;
RA Bateman J.F., Chan D., Walkers I.D., Rogers J.G., Cole W.G.;
RT "Lethal perinatal osteogenesis imperfecta due to the substitution of
RT arginine for glycine at residue 391 of the alpha 1(I) chain of type I
RT collagen.";
RL J. Biol. Chem. 262:7021-7027(1987).
RN [16]
RP VARIANT OI-II CYS-926.
RX MEDLINE=88033031; PubMed=3667599;
RA Vogel B.E., Minor R.R., Freund M., Prockop D.J.;
RT "A point mutation in a type I procollagen gene converts glycine 748
RT of the alpha 1 chain to cysteine and destabilizes the triple helix in
RT a lethal variant of osteogenesis imperfecta.";
RL J. Biol. Chem. 262:14737-14744(1987).
RN [17]
RP VARIANT OI-II ARG-842.
RX MEDLINE=88298828; PubMed=3403550;
RA Bateman J.F., Lamande S.R., Dahl H.H., Chan D., Cole W.G.;
RT "Substitution of arginine for glycine 684 in the collagen alpha 1(I)
RT chain in lethal perinatal osteogenesis imperfecta. Demonstration of
RT the peptide defect by in vitro expression of the mutant cDNA.";
RL J. Biol. Chem. 263:11627-11630(1988).
RN [18]
RP VARIANT OI CYS-1195.
RX MEDLINE=89218828; PubMed=3244312;
RA Labhard M.E., Wirtz M.K., Pope F.M., Nicholls A.C., Hollister D.W.;
RT "A cysteine for glycine substitution at position 1017 in an alpha
RT 1(I) chain of type I collagen in a patient with mild dominantly
RT inherited osteogenesis imperfecta.";
RL Mol. Biol. Med. 5:197-207(1988).
RN [19]
RP VARIANT OI-II VAL-434.
RX MEDLINE=89255493; PubMed=2470760;
RA Patterson E., Smiley E., Boradio J.;
RT "RNA sequence analysis of a perinatal lethal osteogenesis imperfecta
RT mutation.";
RL J. Biol. Chem. 264:10083-10087(1989).
RN [20]
RP VARIANT OI-IV SER-1010.
RX MEDLINE=89308591; PubMed=2745420;
RA Marini J.C., Grange D.K., Gottesman G.S., Lewis M.B., Koeplin D.A.;
RT "Osteogenesis imperfecta type IV. Detection of a point mutation in
RT one alpha 1(I) collagen allele (COL1A1) by RNA/RNA hybrid analysis.";
RL J. Biol. Chem. 264:11893-11900(1989).
RN [21]
RP VARIANTS OI-II ALA-1106, VAL-1151, ARG-1154 AND VAL-1184.
RX MEDLINE=89380165; PubMed=2777764;
RA Lamande S.R., Dahl H.-H.M., Cole W.G., Bateman J.F.;
RT "Characterization of point mutations in the collagen COL1A1 and
RT COL1A2 genes causing lethal perinatal osteogenesis imperfecta.";
RL J. Biol. Chem. 264:15809-15812(1989).
RN [22]
RP VARIANT OI SER-1022.
RX MEDLINE=90062068; PubMed=2511192;
RA Pack M., Constantinou C.D., Kalia K., Nielsen K.B., Prockop D.J.;
RT "Substitution of serine for alpha 1(I)-glycine 844 in a severe
RT variant of osteogenesis imperfecta minimally destabilizes the triple
RT helix of type I procollagen. The effects of glycine substitutions on
RT thermal stability are either position of amino acid specific.";
RL J. Biol. Chem. 264:19694-19699(1989).
RN [23]
RP VARIANT OI-II CYS-1082.
RX MEDLINE=89109573; PubMed=2913053;
RA Constantinou C.D., Nielsen K.B., Prockop D.J.;
RT "A lethal variant of osteogenesis imperfecta has a single base
RT mutation that substitutes cysteine for glycine 904 of the alpha 1(I)
RT chain of type I procollagen. The asymptomatic mother has an
RT unidentified mutation producing an overmodified and unstable type I
RT procollagen.";
RL J. Clin. Invest. 83:574-584(1989).
RN [24]
RP VARIANTS OI CYS-272, CYS-704 AND CYS-896.
RX MEDLINE=90009313; PubMed=2794057;
RA Starman B.J., Eyre D., Charbonneau H., Harrylock M., Weis M.A.,
RA Weiss L., Graham J.M., Byers P.H.;
RT "Osteogenesis imperfecta. The position of substitution for glycine by
RT cysteine in the triple helical domain of the pro alpha 1(I) chains of
RT type I collagen determines the clinical phenotype.";
RL J. Clin. Invest. 84:1206-1214(1989).
RN [25]
RP VARIANT OI-II CYS-422.

QY 1 GSEGGVGRGPPGPPGAGAGPAGDPCGADGPGKAGDAGPAGTAPGPPGARGSPGE 60
 Db 351 GSEGGVGRGPPGPPGAGAGPAGDPCGADGPGKAGDAGPAGTAPGPPGARGSPGE 410
 QY 61 GSGGPPGKGSPPGAGAGPAGDPCGADGPGKAGDAGPAGTAPGPPGARGSPGE 120
 Db 411 GSGGPPGKGSPPGAGAGPAGDPCGADGPGKAGDAGPAGTAPGPPGARGSPGE 470
 QY 121 GPPGGRGPPGPPGAGAGPAGDPCGADGPGKAGDAGPAGTAPGPPGARGSPGE 180
 Db 471 GPPGGRGPPGPPGAGAGPAGDPCGADGPGKAGDAGPAGTAPGPPGARGSPGE 530
 QY 181 GPPGGRGPPGPPGAGAGPAGDPCGADGPGKAGDAGPAGTAPGPPGARGSPGE 240
 Db 531 GPPGGRGPPGPPGAGAGPAGDPCGADGPGKAGDAGPAGTAPGPPGARGSPGE 590
 QY 241 GPPGGRGPPGPPGAGAGPAGDPCGADGPGKAGDAGPAGTAPGPPGARGSPGE 300
 Db 591 GPPGGRGPPGPPGAGAGPAGDPCGADGPGKAGDAGPAGTAPGPPGARGSPGE 650
 QY 301 GPPGGRGPPGPPGAGAGPAGDPCGADGPGKAGDAGPAGTAPGPPGARGSPGE 360
 Db 651 GPPGGRGPPGPPGAGAGPAGDPCGADGPGKAGDAGPAGTAPGPPGARGSPGE 710
 QY 361 GPPGGRGPPGPPGAGAGPAGDPCGADGPGKAGDAGPAGTAPGPPGARGSPGE 420
 Db 711 GPPGGRGPPGPPGAGAGPAGDPCGADGPGKAGDAGPAGTAPGPPGARGSPGE 770
 QY 421 GPPGGRGPPGPPGAGAGPAGDPCGADGPGKAGDAGPAGTAPGPPGARGSPGE 480
 Db 771 GPPGGRGPPGPPGAGAGPAGDPCGADGPGKAGDAGPAGTAPGPPGARGSPGE 830
 QY 481 GPPGGRGPPGPPGAGAGPAGDPCGADGPGKAGDAGPAGTAPGPPGARGSPGE 540
 Db 831 GPPGGRGPPGPPGAGAGPAGDPCGADGPGKAGDAGPAGTAPGPPGARGSPGE 590
 QY 541 GKEG 544
 Db 891 GKZG 894

RESULT 5

CA12 HUMAN STANDARD; PRT; 1418 AA.
 AC P02458
 DT 21-JUN-1986 (Rel. 01, Created)
 DT 01-JAN-1990 (Rel. 13, Last sequence update)
 DT 15-MAR-2004 (Rel. 43, Last annotation update)
 DE Collagen alpha 1(II) chain precursor [Contains: Chondrocalcin].
 GN COL2A1.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=90067946; PubMed=2587267;
 RA Su M.W., Lee B., Ramirez F., Machado M., Horton W.;
 RT "Nucleotide sequence of the full length cDNA encoding for human type
 II procollagen."
 RL Nucleic Acids Res. 17:9473-9473 (1989).
 RN [2]
 RP SEQUENCE OF 1-28 FROM N.A.
 RX MEDLINE=87031574; PubMed=3021582;
 RA Nunez A.M., Kohno K., Martin G.R., Yamada Y.;
 RT "Promoter region of the human pro-alpha 1(II)-collagen gene."
 RL Gene 44:11-16 (1986).
 RN [3]
 RP SEQUENCE OF 432-1145 FROM N.A.
 RX Ramirez F.;
 RA Ramirez F.;
 RL Submitted (DEC-1988) to the EMBL/GenBank/DBJ databases.
 RN [4]

RP SEQUENCE OF 963-1418 FROM N.A.
 RX MEDLINE=85190534; PubMed=3857598;
 RA Cheah K.S.E., Stoker N.G., Griffin J.R., Grosfeld P.G., Solomon E.;
 RT "Identification and characterization of the human type II collagen
 gene (COL2A1)."
 RL Proc. Natl. Acad. Sci. U.S.A. 82:2555-2559 (1985).
 RN [5]
 RP SEQUENCE OF 1120-1398 FROM N.A.
 RX MEDLINE=85306861; PubMed=3840017;
 RA Elima K., Maekelae J.K., Vuorio T., Kauppinen S., Knowles J.,
 Vuorio E.;
 RT "Construction and identification of a cDNA clone for human type II
 procollagen mRNA."
 RL Biochem. J. 229:183-188 (1985).
 RN [6]
 RP SEQUENCE OF 1106-1418 FROM N.A.
 RX MEDLINE=8067771; PubMed=2825137;
 RA Elima K., Vuorio T., Vuorio E.;
 RT "Determination of the single polyadenylation site of the human pro
 alpha 1(II) collagen gene."
 RL Nucleic Acids Res. 15:9499-9504 (1987).
 RN [7]
 RP SEQUENCE OF 1227-1289 FROM N.A.
 RX MEDLINE=86104139; PubMed=3002437;
 RA Nunez A.M., Francomano C., Young M.F., Martin G.R., Yamada Y.;
 RT "Isolation and partial characterization of genomic clones coding for
 a human pro-alpha 1(II) collagen chain and demonstration of
 restriction fragment length polymorphism at the 3' end of the gene."
 RL Biochemistry 24:6343-6348 (1985).
 RN [8]
 RP SEQUENCE OF 1176-1226 FROM N.A.
 RX MEDLINE=84118798; PubMed=6320112;
 RA Strom C.M., Upholt W.B.;
 RT "Isolation and characterization of genomic clones corresponding to
 the human type II procollagen gene."
 RL Nucleic Acids Res. 12:1025-1038 (1984).
 RN [9]
 RP SEQUENCE OF 35-167 FROM N.A.
 RX MEDLINE=89233138; PubMed=2714801;
 RA Su M.W., Benson-Chanda V., Vissing H., Ramirez F.;
 RT "Organization of the exons coding for pro alpha 1(II) collagen N-
 propeptide confirms a distinct evolutionary history of this domain of
 the fibrillar collagen genes."
 RL Genomics 4:438-441 (1989).
 RN [10]
 RP REVIEW ON VARIANTS.
 RX MEDLINE=91184577; PubMed=2010058;
 RA Kuivaniemi H., Tromp G., Prockop D.J.;
 RT "Mutations in collagen genes: causes of rare and some common diseases
 in humans."
 RL FASEB J. 5:2052-2060 (1991).
 RN [11]
 RP REVIEW ON VARIANTS.
 RX MEDLINE=97255959; PubMed=9101290;
 RA Kuivaniemi H., Tromp G., Prockop D.J.;
 RT "Mutations in fibrillar collagens (types I, II, III, and XI), fibril-
 associated collagen (type IX), and network-forming collagen (type X)
 cause a spectrum of diseases of bone, cartilage, and blood vessels."
 RL Hum. Mutat. 9:300-315 (1997).
 RN [12]
 RP VARIANT SER-1074.
 RX MEDLINE=90036909; PubMed=2572591;
 RA Vissing H., D'Alessio M., Lee B., Ramirez F., Godfrey M.,
 Hollister D.W.;
 RT "Glycine to serine substitution in the triple helical domain of pro-
 alpha 1(II) collagen results in a lethal perinatal form of short-
 limbed dwarfism."
 RL J. Biol. Chem. 264:18265-18267 (1989).
 RN [13]
 RP VARIANT SDC 1095-GLY--TYR-1330 DEL.
 RX MEDLINE=89266907; PubMed=2543071;
 RA Lee B., Vissing H., Ramirez F., Rogers D., Romain D.;
 RT "Identification of the molecular defect in a family with

RN spondyloepiphyseal dysplasia.";
 RL Science 244:978-980 (1989).
 RN [14]
 RP VARIANT OSTEOARTHRIITIS CYS-650.
 RX MEDLINE=90370826; PubMed=1975693;
 RA Ala-Kokko L., Baldwin C.T., Moskowitz R.W., Prockop D.J.;
 RT "Single base mutation in the type II procollagen gene (COL2A1) as a
 RT cause of primary osteoarthritis associated with a mild
 RT chondrodysplasia.";
 RL Proc. Natl. Acad. Sci. U.S.A. 87:6565-6568 (1990).
 RN [15]
 RP VARIANT OI-IV VAL-717.
 RX MEDLINE=91291136; PubMed=2064612;
 RA Bateman J.F., Hannagan M., Chan D., Cole W.G.;
 RT "Characterization of a type I collagen alpha 2(I) glycine-586 to
 RT valine substitution in osteogenesis imperfecta type IV. Detection of
 RT the mutation and prenatal diagnosis by a chemical cleavage method.";
 RL Biochem. J. 276:765-770 (1991).
 RN [16]
 RP VARIANT OSTEOARTHRIITIS CYS-650.
 RX MEDLINE=91086471; PubMed=1985108;
 RA Eyre D.R., Weis M.A., Moskowitz R.W.;
 RT "Cartilage expression of a type II collagen mutation in an inherited
 RT form of osteoarthritis associated with a mild chondrodysplasia.";
 RL J. Clin. Invest. 87:357-361 (1991).
 RN [17]
 RP VARIANT HYPOCHONDROGENESIS GLU-984.
 RX MEDLINE=93054548; PubMed=1429602;
 RA Bogaert R., Tiller G.E., Wies M.A., Gruber H.E., Rimoin D.L.,
 RA Cohn D.H., Eyre D.R.;
 RT "An amino acid substitution (Gly853-->Glu) in the collagen alpha
 RT 1(II) chain produces hypochondrogenesis.";
 RL J. Biol. Chem. 267:22522-22526 (1992).
 RN [18]
 RP VARIANT HYPOCHONDROGENESIS SER-705.
 RX MEDLINE=92262484; PubMed=1374906;
 RA Horton W.A., Machado M.A., Ellard J., Campbell D., Bartley J.,
 RA Ramirez F., Vitale E., Lee B.;
 RT "Characterization of a type II collagen gene (COL2A1) mutation
 RT identified in cultured chondrocytes from human hypochondrogenesis.";
 RL Proc. Natl. Acad. Sci. U.S.A. 89:4583-4587 (1992).
 RN [19]
 RP VARIANT WS-II ASP-198.
 RX MEDLINE=93304428; PubMed=8917498;
 RA Koerkoe J., Ritvanemi P., Haataja L., Kaaerlahti H.,
 RA Kivirikko K.I., Prockop D.J., Ala-Kokko L.;
 RT "Mutation in type II procollagen (COL2A1) that substitutes aspartate
 RT for glycine alpha 1-67 and that causes cataracts and retinal
 RT detachment: evidence for molecular heterogeneity in the Wagner
 RT syndrome and the Stickler syndrome (arthro-ophthalmopathy).";
 RL Am. J. Hum. Genet. 53:55-61 (1993).
 RN [20]
 RP VARIANT SEMD CYS-840.
 RA Tiller G.E., Weis M.A., Lachman R.S., Cohn D.H., Rimoin D.L.,
 RA Eyre D.R.;
 RT "A dominant mutation in the type II collagen gene (COL2A1) produces
 RT spondyloepimetaphyseal dysplasia (SEMD), Strudwick type.";
 RL Am. J. Hum. Genet. 53:A209-A209 (1993).
 RN [21]
 RP VARIANT OSTEOARTHRIITIS CYS-650.
 RX MEDLINE=93282819; PubMed=8507190;
 RA Holderbaum D., Maledum C.J., Moskowitz R.W., Haqqi T.M.;
 RT "Human cartilage from late stage familial osteoarthritis transcribes
 RT type II collagen mRNA encoding a cysteine in position 519.";
 RL Biochem. Biophys. Res. Commun. 192:1169-1174 (1993).
 RN [22]
 RP VARIANT SEMD ARG-285.
 RX MEDLINE=93252400; PubMed=8486375;
 RA Vakkula M., Ritvanemi P., Vuorio A.P., Kaitila I., Ala-Kokko L.,
 RA Peltonen L.;
 RT "A mutation in the amino-terminal end of the triple helix of type II
 RT collagen causing severe osteochondrodysplasia.";
 RL Genomics 16:282-285 (1993).

RN [23]
 RP VARIANT SEDC CYS-206.
 RX MEDLINE=94063862; PubMed=8244341;
 RA Williams C.J., Considine E.L., Knowlton R.G., Reginato A., Neumann G.,
 RA Harrison D., Buxton P., Jimenez S.A., Prockop D.J.;
 RT "Spondyloepiphyseal dysplasia and precocious osteoarthritis in a
 RT family with an Arg75-->Cys mutation in the procollagen type II gene
 RT (COL2A1).";
 RL Hum. Genet. 92:499-505 (1993).
 RN [24]
 RP VARIANT SEDC CYS-920.
 RX MEDLINE=93315508; PubMed=8325895;
 RA Chan D., Taylor T.K.F., Cole W.G.;
 RT "Characterization of an arginine 789 to cysteine substitution in
 RT alpha 1 (II) collagen chains of a patient with spondyloepiphyseal
 RT dysplasia.";
 RL J. Biol. Chem. 268:15238-15245 (1993).
 RN [25]
 RP VARIANT SEDC SER-1128.
 RX MEDLINE=93140139; PubMed=8423604;
 RA Cole W.G., Hall R.K., Rogers J.G.;
 RT "The clinical features of spondyloepiphyseal dysplasia congenita
 RT resulting from the substitution of glycine 997 by serine in the alpha
 RT 1(II) chain of type II collagen.";
 RL J. Med. Genet. 30:27-35 (1993).
 / Query Match 71.7%; Score 2202; DB 1; Length 1418;
 Best Local Similarity 71.9%; Pred. No. 1.3e-82;
 Matches 391; Conservative 40; Mismatches 113; Indels 0; Gaps 0;
 QY 1 GSEGGVGVGPGPGPAGAGPAGDPGADGPGAGKAGDAGPAGIAGAPGPGARGPSGPE 60
 Db 315 GREGAGGPRGPGTQSPGAPAGASGNFTDGI PGAGKSAGAGIAGAPGPGRPDPQ 374
 QY 61 GPGGPGPKGDSGEPGAPGSKGDTAKGPPGVGVGPPGPGAGEGKPGARPGTGLP 120
 Db 375 GATGLPGKGTCKPGIAGFKGEPKPGPPGAGPAGPAGPAGGEGKRGARGEPGVGP 434
 QY 121 GPPGEGGPGSGPFGADGVAGPKGAPGAGSPGAGPKGSGPAGPAGPAGKGLT 180
 Db 435 GPPGEGGPGSGPFGADGVAGPKGAPGAGSPGAGPKGSGPAGPAGPAGKGLT 494
 QY 181 GSPGSPGPGDKTGPFGPAGDGRPGPPGPGARGGAGVGVGPPGKGAAGPAGRGVP 240
 Db 495 GRPGDAGPQGVGSPGAPGCDGDRPGPPGPGAGQGVGPPGKANGEPKAGEKGLP 554
 QY 241 GPPGAVGPGAGKDEAGAPGPPGAGPAGGEGGPGAGSPGFBGLPGPAGPAGKPGEE 300
 Db 555 GAFGLRGLPGKDGETGAEPPGPGPAGPAGRGEGGAPGSGFQGLPGPPGPGEGKPGDQ 614
 QY 301 GVPDILGAPGSPGARGPAGPAGPAGPAGPAGPAGPAGPAGPAGPAGPAGPAG 360
 Db 615 GVPGEAGAPGLVGRGPRGPRGPRGPRGPRGPRGPRGPRGPRGPRGPRGPRG 674
 QY 361 GAFGLEMPGEGAGAGLPKPGDRGAGPKGADGSPKDGVRGLTGPIGPPGAGAGDK 420
 Db 675 GPPGLQMPGEGAGAGTAGPKGDRGAGPAGPAGPAGPAGPAGPAGPAGPAGPAG 734
 QY 421 GEGSPGSPGAGTARGAPGDRGPPGPPGAGPAGPAGPAGPAGPAGPAGPAGPAG 480
 Db 735 GEVGPAGPAGSAGARGAPRGRTGTPGTSIAGPAGAGQPGKAGEAGKQKAGAP 794
 QY 481 GPAGPAGPAGPAGPAGPAGPAGPAGPAGPAGPAGPAGPAGPAGPAGPAGPAG 540
 Db 795 GPQGPSGAPGPGTGTGTGPKARGAQGPPGATGTFPAGAGRVGPPGPPGPPGPP 854
 QY 541 GKEG 544
 Db 855 KDG 858

RESULT 6
 CA12_MOUSE

ID CA12 MOUSE STANDARD; PRT; 1459 AA.
 AC P28481;
 DT 01-DEC-1992 (Rel. 24, Created)
 DT 01-DEC-1992 (Rel. 24, Last sequence update)
 DE 15-MAR-2004 (Rel. 43, Last annotation update)
 GN Collagen alpha 1(II) chain precursor [Contains: Chondrocalcin].
 GN COL2A1.
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 OX NCBI_TaxID=10090;
 RN [1]
 RP SEQUENCE FROM N.A., AND ALTERNATIVE SPLICING.
 RX MEDLINE=91358489; PubMed=1885613;
 RA Metsaranta M., Toman D., de Crombrughe B., Vuorio E.;
 RT "Mouse type II collagen gene. Complete nucleotide sequence, exon
 structure, and alternative splicing.";
 RL J. Biol. Chem. 266:16862-16869 (1991).
 RN [2]
 RP SEQUENCE OF 1455-1459 FROM N.A.
 RX MEDLINE=91274355; PubMed=2054384;
 RA Metsaranta M., Toman D., de Crombrughe B., Vuorio E.;
 RT "Specific hybridization probes for mouse type I, II, III and IX
 collagen mRNAs.";
 RL Biochim. Biophys. Acta 1089:241-243 (1991).
 CC -!- FUNCTION: Collagen type II is specific for cartilaginous tissues.
 CC -!- SUBUNIT: Homotrimers of alpha 1(II) chains.
 CC -!- ALTERNATIVE PRODUCTS:
 CC Event=Alternative splicing; Named isoforms=2;
 CC Name=Long;
 CC IsoId=28481-1; Sequence=Displayed;
 CC Name=Short;
 CC IsoId=28481-2; Sequence=VSP_001139, VSP_001140;
 CC -!- PM: Prolines at the third position of the tripeptide repeating
 CC unit (G-X-Y) are hydroxylated in some or all of the chains.
 CC -!- SIMILARITY: Contains 1 WFC domain.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL; M65161; AAA68100.1; -;
 DR EMBL; X57982; CAA41047.1; -;
 DR MGD; MGI:88452; Col2a1.
 DR InterPro; IPR008161; Clq helix.
 DR InterPro; IPR008160; Collagen.
 DR InterPro; IPR000885; Fib_collagen_C.
 DR InterPro; IPR001007; WVF_C.
 DR Pfam; PF01410; COLFI; 1.
 DR Pfam; PF01391; Collagen; 18.
 DR Pfam; PF00093; wvc; 1.
 DR ProDom; PD000007; Clq helix; 5.
 DR ProDom; PD002078; Fib_collagen_C; 1.
 DR SMART; SMO0038; COLFI_1.
 DR SMART; SMO0214; WVC; 1.
 DR PROSITE; PS01208; WVF_1; 1.
 DR PROSITE; PS50184; WVF_C; 1.
 KW Extracellular matrix; Connective tissue; Repeat; Hydroxylation;
 KW Glycoprotein; Collagen; Signal; Alternative splicing.
 FT SIGNAL 1 25 POTENTIAL.
 FT PROPEP 26 153 AMINO-TERMINAL PROPEPTIDE
 FT (BY SIMILARITY).
 FT CHAIN 154 1213 COLLAGEN ALPHA 1(II) CHAIN.
 FT PROPEP 1214 1459 CHONDROCALCIN.
 FT DOMAIN 32 89 WVC.
 FT DOMAIN 173 1186 TRIPLE-HELICAL REGION.
 FT DOMAIN 1187 1213 NONHELICAL REGION (C-TERMINAL).
 FT VARSP LIC 29 29 Q -> R (in isoform Short).
 FT /FTid=VSP_001139.

FT VARSP LIC 30 98 Missing (in isoform Short).
 FT /FTid=VSP_001140.
 SQ SEQUENCE 1459 AA; 139154 MW; F6C84FA7C532E7F2 CRC64;
 Query Match 71.4%; Score 2192; DB 1; Length 1459;
 Best Local Similarity 71.7%; Pred. No. 3.4e-82;
 Matches 390; Conservative 40; Mismatches 114; Indels 0; Gaps 0;
 QY 1 GSEPGVNRGEPPGPPGAGAGPAGDPADGDPFGAGKADGAGPAGIAGAGPFGGARGPSGPE 60
 DB 356 GPEGAGQSRGEPNPGSPGAGAGNPGTDGIPGAKGSAGAPGAGPFGPGRGPPGQ 415
 QY 61 GGGPGPGKDSGEPCAPGSKGDTGAKGEPGVGVGPPGPRGPRGPRGPRGPRGPRGPRG 120
 DB 416 GATGPGKPGQAGEPGAGIAGFGKDGQKQKGTGTPAGPQAGPAGGEGKRGARGEPGAGPI 475
 QY 121 GPPGPGGSGRGTGADGVAGKPGPAGERSGPPGAGKSPGSPGSPGSPGSPGSPGSPGSPG 180
 DB 476 GPPGPGGAPNKGFTGQDGLAGKPGKAPGERSGSLAGPKGANGDGRPGPCLPGARGLT 535
 QY 181 GSPGPGPDGKTGPPGPPAGEDDGRPPGPPGARGAGVGMPPGPKGAAGEPGKAGRGVP 240
 DB 536 GPPGPGGKGVGSGAPGDEGRPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPG 595
 QY 241 GPPGAVGPAKDGAGAGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPG 300
 DB 596 GAGPLRGLPGKDGETGAGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPG 655
 QY 301 GVPDGLGAPSGARGEPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPG 360
 DB 656 GPPGAGAPGLVPRGERGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPG 715
 QY 361 GAGLEGMPGPRGAGLPGKDRDAGKDGADSGPKDGVRLTGRTPGPPGPPGPPGPPG 420
 DB 716 GPPGQGVGPRGAGIAGPKGDRGDKGKPEAGPKDGGRLTGRTPGPPGPPGPPG 775
 QY 421 GSEPGSPAGTGAAGAPGDRGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPP 480
 DB 776 GEGAPGPPSGSTGARGAPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPG 835
 QY 481 GPAGAGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPP 540
 DB 836 GPPGSPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPP 895
 QY 541 GKEG 544
 DB 896 GKDG 899
 RESULT 7
 CALL RAT
 ID CALL RAT STANDARD; PRT; 671 AA.
 AC P02454; P02455;
 DT 21-JUL-1986 (Rel. 01, Created)
 DT 01-FEB-1994 (Rel. 28, Last sequence update)
 DT 28-FEB-2003 (Rel. 41, Last annotation update)
 DE Collagen alpha 1(I) chain (fragments).
 GN COL1A1.
 OS Rattus norvegicus (Rat).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
 OX NCBI_TaxID=10116;
 RN [1]
 RP SEQUENCE OF 1-19.
 RX MEDLINE=69155173; PubMed=5777344;
 RA Bornstein P.;
 RT "Comparative sequence studies of rat skin and tendon collagen. II.
 RT The absence of a short sequence at the amino terminus of the skin
 RT alpha-1 chain.";
 RL Biochemistry 8:63-71 (1969).
 RN [2]
 RP SEQUENCE OF 5-19.
 RX MEDLINE=67162268; PubMed=5337886;

RA Kang A.H., Bornstein P., Piez K.A.;
 RT "The amino acid sequence of peptides from the cross-linking region of
 RT rat skin collagen.";
 RL Biochemistry 6:788-795(1967).
 RN [3]
 RP SEQUENCE OF 20-55.
 RA MEDLINE=67165368; PubMed=4290711;
 RA Bornstein P.;
 RT "The incomplete hydroxylation of individual prolyl residues in
 RT collagen.";
 RL J. Biol. Chem. 242:2572-2574(1967).
 RN [4]
 RP SEQUENCE OF 56-102.
 RA MEDLINE=71263178; PubMed=4327399;
 RA Butler W.T., Ponds S.L.;
 RT "Chemical studies on the cyanogen bromide peptides of rat skin
 RT collagen. Amino acid sequence of alpha 1-CB4.";
 RL Biochemistry 10:2076-2081(1971).
 RN [5]
 RP SEQUENCE OF 103-139.
 RA MEDLINE=70085124; PubMed=5411206;
 RA Butler W.T.;
 RT "Chemical studies on the cyanogen bromide peptides of rat skin
 RT collagen. The covalent structure of alpha 1-CB5, the major
 RT hexose-containing cyanogen bromide peptide of alpha 1.";
 RL Biochemistry 9:44-50(1970).
 RN [6]
 RP SEQUENCE OF 140-238.
 RA MEDLINE=72136131; PubMed=4335087;
 RA Balian G., Click E.M., Bornstein P.;
 RT "Structure of rat skin collagen alpha 1-CB8. Amino acid sequence of
 RT the hydroxylamine-produced fragment HA1.";
 RL Biochemistry 10:4470-4478(1971).
 RN [7]
 RP SEQUENCE OF 239-418.
 RA MEDLINE=73006942; PubMed=4342027;
 RA Balian G., Click E.M., Hermodson M.A., Bornstein P.;
 RT "Structure of rat skin collagen alpha 1-CBB. Amino acid sequence of
 RT the hydroxylamine-produced fragment HA2.";
 RL Biochemistry 11:3798-3806(1972).
 RN [8]
 RP SEQUENCE OF 419-567.
 RA MEDLINE=74271984; PubMed=4366532;
 RA Butler W.T., Underwood S.P., Finch J.E. Jr.;
 RT "Chemical studies on the cyanogen bromide peptides of rat skin
 RT collagen. Amino acid sequence of alpha 1-CB3.";
 RL Biochemistry 13:2946-2953(1974).
 RN [9]
 RP SEQUENCE OF 568-651.
 RA MEDLINE=74011954; PubMed=4126850;
 RA Stoltz M., Timpl R., Furthmayr H., Kuehn K.;
 RT "Structural and immunogenic properties of a major antigenic
 RT determinant in neutral salt-extracted rat-skin collagen.";
 RL Eur. J. Biochem. 37:287-294(1973).
 RN [10]
 RP SEQUENCE OF 651-671.
 RA MEDLINE=73049495; PubMed=4636751;
 RA Stoltz M., Timpl R., Kuehn K.;
 RT "Non-helical regions in rat collagen alpha 1-chain.";
 RL FEBS Lett. 26:61-65(1972).
 RN [11]
 RP SEQUENCE OF 529-567 FROM N.A.
 RA MEDLINE=85122694; PubMed=6395893;
 RA Genovese C., Rowe D., Kream B.;
 RT "Construction of DNA sequences complementary to rat alpha 1 and alpha
 RT 2 collagen mRNA and their use in studying the regulation of type I
 RT collagen synthesis by 1,25-dihydroxyvitamin D.";
 RL Biochemistry 23:6210-6216(1984).
 CC -!- FUNCTION: Type I collagen is a member of group I collagen
 CC (fibrillar forming collagen).
 CC -!- SUBUNIT: Trimers of one alpha 2(I) and two alpha 1(I) chains.
 CC -!- TISSUE SPECIFICITY: Forms the fibrils of tendon, ligaments and
 CC bones. In bones the fibrils are mineralized with calcium

hydroxyapatite.
 CC -!- PTM: Proline residues at the third position of the tripeptide
 CC repeating unit (G-X-Y) are hydroxylated in some or all of the
 CC chains.
 CC -!- PTM: O-linked glycan consists of a Glc-Gal disaccharide bound to
 CC the oxygen atom of a post-translationally added hydroxyl group.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL collaboration
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL; M14432; AAA40832.1; ALT SEQ.
 DR InterPro; IPR008161; C1g helix.
 DR InterPro; IPR008160; Collagen.
 DR InterPro; IPR001007; VWF C.
 DR ProDom; PD000007; C1g_helix; 1.
 DR PROSITE; PS01208; VWF_C1; PARTIAL.
 KW Extracellular matrix; Connective tissue; Repeat; Hydroxylation;
 KW Glycoprotein; Collagen; Pyrrolidone carboxylic acid.
 FT MOD_RES 1 1 PYRROLIDONE CARBOXYLIC ACID (PROBABLE).
 FT MOD_RES 9 9 CONVERTED TO AN ALDEHYDE GROUP THAT IS
 FT INVOLVED IN CROSS-LINKING.
 FT MOD_RES 28 28 HYDROXYLATION (PROBABLE).
 FT MOD_RES 31 31 HYDROXYLATION (PROBABLE).
 FT MOD_RES 34 34 HYDROXYLATION (PROBABLE).
 FT MOD_RES 43 43 HYDROXYLATION (PROBABLE).
 FT MOD_RES 46 46 HYDROXYLATION (PROBABLE).
 FT MOD_RES 49 49 HYDROXYLATION (PROBABLE).
 FT MOD_RES 103 103 O-LINKED (GAL..).
 FT CARBOHYD 103 103 HYDROXYLATION (PROBABLE).
 FT MOD_RES 424 424 HYDROXYLATION (PROBABLE).
 FT MOD_RES 547 547
 FT NON_CONS 567 568
 FT DOMAIN 641 651
 FT MAJOR ANTIGENIC DETERMINANT (OF NEUTRAL
 FT SALT-EXTRACTED RAT SKIN COLLAGEN).
 SQ SEQUENCE 671 AA; 60615 MW; 9DC3114204AC4918 CRC64;
 Query Match 70.1%; Score 2151; DB 1; Length 671;
 Best Local Similarity 77.5%; Pred. No. 9.1e-81;
 Matches 382; Conservative 27; Mismatches 36; Indels 48; Gaps 2;
 QY 1 GSEGGVGEVGEPPGPPGAGAGPAGDPGADGEPGAKGADGAPGIAGAPGPPGARGPSGE 60
 DB 200 GSEGGVGEVGEPPGPPGAGAGPAGDPGADGEPGAKGADGAPGIAGAPGPPGARGPSGQ 259
 QY 61 GPGGPPGKGDSEPGAPGSKGDTGAKGPPGVGVGPPGPPGAGEGKPGARGEPGTGLP 120
 DB 260 GPSGAPGPKGNSGEPGAPGNKGDGAKGPPGAGVQVQPPGAGEGKRGARGEPGSGLP 319
 QY 121 GPPGERGPGSRGFPAGDCVAGPKGPPAGRGSGPPGAPGKSGPGEAGRPCEAGLPGAKGLT 180
 DB 320 GPPGERGPGSRGFPAGDCVAGPKGPPAGRGSGPPGAPGKSGPGEAGRPCEAGLPGAKGLT 379
 QY 181 GSPGSPGPDGKTGPPGPPGAGEDGPPGPPGARGAGVGMVFFGPKGAAGEPGKAGERGVP 240
 DB 380 GSPGSPGPDGKTGPPGPPGAGEDGPPGPPGARGAGVGMVFFGPKGTACEPGKAGERGVP 439
 QY 241 GPPGAVGPAGKDGAEAGPPGPPGAGERGEGGAGSGPFGELPGPAGPPEAGKPGEE 300
 DB 440 GPPGAVGPAGKDGAEAGPPGPPGAGERGEGGAGSGPFGELPGPAGPPEAGKPGGZ 499
 QY 301 GVPEDLGAPGSPGARGEPGPPGPPGAGEGPPGPPGAGDGDGAKGADGAPGAPGSE 360
 DB 500 GVPEDLGAPGSPGARGEPGPPGPPGAGEGPPGPPGAGPRGNGAGDNGAKGDTGAPGAPGSG 559
 QY 361 GAPGLGMPGEPGAGLPGPKGDRGAPGKADGSPGKDGVRGLTGPIPPPPGAPGAPGK 420
 DB 560 GAPGLGMPGSLZ-----GPPGPPGSPGZ 583


```

QY 532 GPPGPPGAGKGG 544
Db 899 GPPGHPGAGNG 911

RESULT 11
CA21_RAT
ID CA21_RAT STANDARD; PRT; 1372 AA.
AC P02466; Q9R188;
DT 21-JUL-1986 (Rel. 01, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Collagen alpha 2(I) chain precursor.
GN COL1A2.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_taxid=10116;
RN [1]
RP SEQUENCE FROM N.A.
RA Guenther D., Seibold S., Marx M.;
RL Submitted (JAN-1999) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE OF 86-98.
RC TISSUE=Skin;
RX MEDLINE=67162268; PubMed=5337886;
RA Kang A.H., Bornstein P., Piez K.A.;
RT "The amino acid sequence of peptides from the cross-linking region of
RT rat skin collagen.";
RL Biochemistry 6:788-795 (1967).
RN [3]
RP SEQUENCE OF 99-102.
RC TISSUE=Skin;
RX MEDLINE=69206881; PubMed=5785232;
RA Fietzek P.P., Piez K.A.;
RT "Isolation and characterization of the cyanogen bromide peptides from
RT the alpha 2 chain of rat skin collagen.";
RL Biochemistry 8:2129-2133 (1969).
RN [4]
RP SEQUENCE OF 102-144.
RC TISSUE=Skin;
RX MEDLINE=73049496; PubMed=4636752;
RA Fietzek P.P., Kell I., Kuehn K.;
RT "The covalent structure of collagen. Amino acid sequence of the N-
RT terminal region of alpha 2-CB4 from calf and rat skin collagen.";
RL FEBS Lett. 26:66-68 (1972).
RN [5]
RP SEQUENCE OF 423-452.
RC TISSUE=Skin;
RX MEDLINE=71115216; PubMed=5544553;
RA Hightberger J.H., Kang A.H., Gross J.;
RT "Comparative studies on the amino acid sequence of the alpha 2-CB2
RT peptides from chick and rat skin collagens.";
RL Biochemistry 10:610-616 (1971).
RN [6]
RP SEQUENCE OF 453-501.
RC TISSUE=Skin;
RX MEDLINE=75059250; PubMed=4435743;
RA Fietzek P.P., Kuehn K.;
RT "The covalent structure of collagen: amino acid sequence of the N-
RT terminal region of alpha2-CB3 from rat skin collagen and alpha2-CB3.5
RT from calf skin collagen.";
RL Hoppe-Seyler's Z. Physiol. Chem. 355:647-650 (1974).
RN [7]
RP SEQUENCE OF 791-836.
RC TISSUE=Skin;
RX MEDLINE=74055004; PubMed=4763308;
RA Fietzek P.P., Kuehn K.;
RT "The covalent structure of collagen: amino acid sequence of the N-
RT terminal region of alpha 2-CB5 from rat skin collagen.";
RL FEBS Lett. 36:289-291 (1973).
RN [8]

```

```

RP ORDER OF CNBR PEPTIDES.
RX MEDLINE=70181852; PubMed=5443712;
RA Vuust J., Lane J.M., Fietzek P.P., Miller E.J., Piez K.A.;
RT "The order of the CNBR peptides from the alpha 2 chain of collagen.";
RL Biochem. Biophys. Res. Commun. 38:703-708 (1970).
CC -!- FUNCTION: Type I collagen is a member of group I collagen
CC (fibrillar forming collagen).
CC -!- SUBUNIT: Trimers of one alpha 2(I) and two alpha 1(I) chains.
CC -!- TISSUE SPECIFICITY: Forms the fibrils of tendon, ligaments and
CC bones. In bones the fibrils are mineralized with calcium
CC hydroxyapatite.
CC -!- PTM: Prolines at the third position of the tripeptide repeating
CC unit (G-X-Y) are hydroxylated in some or all of the chains.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: AF121217; AAD41775.1; -.
DR InterPro: IPR008161; Clg_helix.
DR InterPro: IPR008160; Collagen.
DR InterPro: IPR000885; Fib_collagen_C.
DR Pfam: PF01410; COLFI; 1.
DR Pfam: PF01391; Collagen; 18.
DR ProDom: PD000007; Clg_helix; 4.
DR ProDom: PD002078; Fib_collagen_C; 1.
DR SMART: SM00038; COLFI; 1.
DR Extracellular matrix; Connective tissue; Repeat; Hydroxylation;
KW Glycoprotein; Collagen; Signal; Pyrrolidone carboxylic acid.
KW SIGNAL 1 24 POTENTIAL.
FT PROPEP 25 85 AMINO-TERMINAL PROPEPTIDE
FT CHAIN 86 1108 COLLAGEN ALPHA 2(I) CHAIN
FT PROPEP 1109 1372 CARBOXYL-TERMINAL PROPEPTIDE
FT (BY SIMILARITY).
FT SITE 783 785 CELL ATTACHMENT SITE (POTENTIAL).
FT SITE 828 830 CELL ATTACHMENT SITE (POTENTIAL).
FT SITE 1011 1013 CELL ATTACHMENT SITE (POTENTIAL).
FT MOD_RES 86 86 PYRROLIDONE CARBOXYLIC ACID (PROBABLE).
FT MOD_RES 90 90 CONVERTED TO AN ALDEHYDE GROUP THAT IS
FT INVOLVED IN CROSS-LINKING.
FT CARBOHYD 1273 1273 N-LINKED (GLCNAC...) (POTENTIAL).
FT CONFLICT 132 132 T -> P (IN REF. 4).
FT CONFLICT 137 137 S -> P (IN REF. 4).
FT CONFLICT 145 422 MISSING (IN REF. 4).
FT CONFLICT 431 432 ST -> TS (IN REF. 5).
FT CONFLICT 494 494 E -> Z (IN REF. 6).
FT CONFLICT 497 497 N -> A (IN REF. 6).
FT CONFLICT 502 790 MISSING (IN REF. 6).
FT CONFLICT 825 825 R -> K (IN REF. 7).
SQ SEQUENCE 1372 AA; 129564 MW; B069371A8DB20A72 CRC64;

Query Match 63.0%; Score 1934; DB 1; Length 1372;
Best Local Similarity 64.7%; Pred. No. 7,8e-72;
Matches 352; Conservative 44; Mismatches 148; Indels 0; Gaps 0;

QY 1 GSEGGVGRGPPGPPGAGAGPAGDPGADGPGAGKAGDAGPAGIAGAGPFGARGSPGE 60
Db 280 GPAGPAGPGEAGLPGLSGPVGPPGNGPANGLTGAKGATGLPGVAGAPGLGPRGIPGV 339
QY 61 GPGGPPGKDSGEPGAPGSKGDTGAKGPPGVGVGPPGPPGAGEKPGARGEPGTGLP 120
Db 340 GAAGATGPRGLVGBEPGAPGSKGTGNKGPPGSAGAQGPPGSGGKRGSPGSAGPA 399
QY 121 GPPGGRGSGRGFPAGADGAVAGFKGPPAGRGSPGAGPKGSGPGEAGLPKAGLT 180
Db 400 GPPGLRGSGRGLPGADGAGVMPGPPGNRGSTGAGVGGPNDAGRGPPGLMGRGLP 459
QY 181 GSPGSPGDPGKTPGPPGAGEDGRPGPPPPGARGAGVMPGPPGKGAAGEPGKAGERGV 240

```

```

Db 460 GSPGNVPAGKEGVGLFIDGRPGTGPAGPRGEAGNIGFPQPKPSGDPGKPGSKGHP 519
Qy 241 GPPGAVGAPGAKDAGAGEGPPGAPGAGGEGEPGAGSPGFEGLPGPAGPPGAGKPGEE 300
Db 520 GLAGARGAPGPDGNNAGQGGPPGVGVQGGKGGGAGPPGFGQLPGSPGTAGEVGFGER 579
Qy 301 GVPGLDGLAPSGARGPSPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPG 360
Db 580 GLPGEFLGPPGAPGRGERPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPG 639
Qy 361 GAPLEGWPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPG 420
Db 640 GASPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPP 699
Qy 421 GESPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPP 480
Db 700 GEAGAAGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPP 759
Qy 481 GPAGPAGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPP 540
Db 760 GPTGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPG 819
Qy 541 GKEG 544
Db 820 GKEG 823

RESULT 12
CA25 HUMAN STANDARD; PRT; 1496 AA.
ID -CA25 HUMAN STANDARD; PRT; 1496 AA.
AC P05997;
DC 01-APR-1988 (Rel. 07, Created)
DT 01-JAN-1990 (Rel. 13, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Collagen alpha 2(V) chain precursor.
GN COL5A2.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE OF 1-463 FROM N.A.
RX MEDLINE=89123368; PubMed=2914927;
RA Woodbury D., Benson-Chanda V., Ramirez F.;
RT "Amino-terminal propeptide of human pro-alpha 2(V) collagen conforms
RT to the structural criteria of a fibrillar procollagen molecule.";
RL J. Biol. Chem. 264:2735-2738(1989).
RN [2]
RP SEQUENCE OF 398-1496 FROM N.A.
RX MEDLINE=87146331; PubMed=3029669;
RA Weil D., Bernard M.P., Gargano S., Ramirez F.;
RT "The pro alpha 2(V) collagen gene is evolutionarily related to the
RT major fibrillar-forming collagens.";
RL Nucleic Acids Res. 15:181-198(1987).
RN [3]
RP SEQUENCE OF 1227-1496 FROM N.A.
RX MEDLINE=85289337; PubMed=2411731;
RA Myers J.C., Loidl H.R., Seyer J.M., Dion A.S.;
RT "Complete primary structure of the human alpha 2 type V procollagen
RT COOH-terminal propeptide.";
RL J. Biol. Chem. 260:11216-11222(1985).
RN [4]
RP SEQUENCE OF 1449-1496 FROM N.A.
RX MEDLINE=89138450; PubMed=3224983;
RA Tsipouras P., Schwartz R.C., Liddell A.C., Salkeld C.S., Weil D.,
RA Ramirez F.;
RT "Genetic distance of two fibrillar collagen loci, COL3A1 and COL5A2,
RT located on the long arm of human chromosome 2.";
RL Genomics 3:275-277(1988).
RN [5]
RP SEQUENCE OF 208-227.
RC TISSUE=Placenta;

```

```

RX MEDLINE=922339022; PubMed=1571108;
RA Mann K.;
RT "Isolation of the alpha 3-chain of human type V collagen and
RT characterization by partial sequencing.";
RL Biol. Chem. Hoppe-Seyler 373:69-75(1992).
RN [6]
RP SEQUENCE OF 288-297 AND 606-617.
RC TISSUE=Bone;
RX MEDLINE=94237164; PubMed=8181482;
RA Moradi-Ameli M., Rousseau J.C., Kleman J.P., Champliand M.F.,
RA Boutillon M.M., Bernillon J., Wallach J.M., van der Rest M.;
RT "Diversity in the processing events at the N-terminus of type-V
RT collagen.";
RL Eur. J. Biochem. 221:987-995(1994).
RN [7]
RP DISEASE.
RX MEDLINE=98087576; PubMed=9425231;
RA Michalikova K., Susic M., Willing M.C., Wenstrup R.J., Cole W.G.;
RT "Mutations of the alpha2(V) chain of type V collagen impair matrix
RT assembly and produce Ehlers-Danlos syndrome type I.";
RL Hum. Mol. Genet. 7:249-255(1998).
RN [8]
RP VARIANT EDS-II ARG-960.
RX MEDLINE=98455031; PubMed=9783710;
RA Richards A.J., Martin S., Nicholls A.C., Harrison J.B., Pope F.M.,
RA Burrows N.P.;
RT "A single base mutation in COL5A2 causes Ehlers-Danlos syndrome type
RT II.";
RL J. Med. Genet. 35:846-848(1998).
CC -!- FUNCTION: Type V collagen is a member of group I collagen
CC (fibrillar forming collagen). It is a minor connective tissue
CC component of nearly ubiquitous distribution. Type V collagen binds
CC to DNA, heparan sulfate, thrombospondin, heparin, and insulin.
CC -!- SUBUNIT: Trimers of two alpha 1(V) and one alpha 2(V) chains in
CC most tissues and trimers of one alpha 1(V), one alpha 2(V), and
CC one alpha 3(V) chains in placenta.
CC -!- PTM: Prolines at the third position of the tripeptide repeating
CC unit (G-X-Y) are hydroxylated in some or all of the chains.
CC -!- DISEASE: Defects in COL5A2 are a cause of Ehlers-Danlos syndrome
CC type I (EDS-I) [MIM:130000]; also known as Ehlers-Danlos syndrome
CC gravis. EDS-I is an autosomal dominant connective-tissue disorder
CC characterized by loose-jointedness and fragile, velvety,
CC stretchable, bruisable skin that heals with peculiar 'cigarette-
CC paper' scars. Inheritance is autosomal dominant.
CC -!- DISEASE: Defects in COL5A2 are a cause of Ehlers-Danlos syndrome
CC type II (EDS-II) [MIM:130010]; also known as Ehlers-Danlos
CC syndrome mitis. Inheritance is autosomal dominant.
CC -!- SIMILARITY: Contains 1 WFSC domain.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; J04478; AAA51859.1; -
CC EMBL; X04758; CAA28454.1; -
CC EMBL; M11718; AAA52058.1; -
CC PIR; A31427; CGHU2V.
CC PDB; 1A9A; 18-NOV-98.
CC Genew; HGNC:2210; COL5A2.
CC MIM; 120190; -
CC MIM; 130000; -
CC MIM; 130010; -
CC GO; GO:0005588; C:collagen type V; TAS.
CC GO; GO:0008151; P:cell growth and/or maintenance; TAS.
CC InterPro; IPR008161; Clg helix.
CC InterPro; IPR008160; Collagen.
CC InterPro; IPR000885; Fib collagen C.
CC InterPro; IPR002181; Fibrinogen_C.
CC InterPro; IPR001007; VWF_C.

```

```

DR Pfam; PF01410; COLFI; 1.
DR Pfam; PF01391; Collagen; 18.
DR Pfam; PF00093; vwc; 1.
DR ProDom; PD000007; C1g_helix; 5.
DR ProDom; PD002078; Fib_collagen_C; 1.
DR SMART; SM00038; COLFI; 1.
DR SMART; SM00214; vwc; 1.
DR PROSITE; PS01208; VWC_1; 1.
DR PROSITE; PS01184; VWC_2; 1.
KW Extracellular matrix; Connective tissue; Repeat; Hydroxylation;
KW Glycoprotein; Collagen; Signal; Ehlers-Danlos syndrome;
KW Disease mutation; 3D-structure.
FT SIGNAL 1 26
FT CHAIN 27 1226
FT PROPEP 1227 1496
FT DOVAIN 39 97
FT MOD_RES 230 290
FT MOD_RES 293 293
FT MOD_RES 296 296
FT MOD_RES 608 614
FT MOD_RES 614 614
FT VARIANT 960 960
FT CONFLICT 292 292
FT CONFLICT 1418 1418
FT CONFLICT 1438 1438
FT CONFLICT 1460 1460
FT CONFLICT 1496 1496
SQ SEQUENCE 1496 AA; 144720 MW; 82827C17A8644F5A CRC64;

Query Match 62.6%; Score 1921.5; DB 1; Length 1496;
Best Local Similarity 64.0%; Pred. No. 2.6e-71;
Matches 348; Conservative 40; Mismatches 153; Indels 3; Gaps 1;

QY 1 CSEGGVGRGPPGPPGAGAGADPCADCEPCAKGADGAPGAGPPGARGPSGPE 60
Db
396 GPESQGGRTGTPGPPGVSGPLGATGDTTPPKPTGSGTSGP---PGSAGPPSP 452
QY 61 GPGGPPKGGSGPAGPSKXGDTGAKGPPGVGVGPPGPPGAGBEGKPGARFPGTGLP 120
Db
453 GPQSGTGQSGNSGLPGDPFGKGEAGKPGEPHGIQPGIPGPEGKRGPRGDPDTLGP 512
QY 121 GPPGEGGSGRGPAGDVAGPKGAPGERSGPGAPKSGPSGAGRGAGLPGAKGLT 180
Db
513 GPVGERGAPGNKGFSGDGLPFGKAGQGERGPPVSSGPKSGQDPPGPPGFLPGARGLT 572
QY 181 GSPGSGPDGKTGTPPGAGEDCRPGPPGPPGARGGAGVMPGPPGKGAAGPPGKAGERGVP 240
Db
573 GNPVGQGPGLGPGAGDEDRPGPSIGIKQPGTGLPGPKGNGDPKRGAGNP 632
QY 241 GPPGAVGPPAGKDGAGAGPPGPPGAPGAGERGEPGAGSPGPEGLPPGPPGAGKGE 300
Db
633 GVPQORGAPGKDGKVPYGPFGPPGLRGERGEGEQPGPTGFGQHPGPPGPPGEGKPGDQ 692
QY 301 GVPDGLGAPGSGARGEPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPG 360
Db
693 GVPFGGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPG 752
QY 361 GAPLEGMPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPG 420
Db
753 GPPGLQMPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPG 812
QY 421 GESGSPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPG 480
Db
813 GEPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPG 872
QY 481 GPAGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPG 540
Db
873 GPQGLAGSPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPP 932
QY 541 GKEG 544
Db
933 GKEG 936

```

RESULT 13

```

CA21 MOUSE
ID -CA21 MOUSE STANDARD; PRT; 1372 AA.
AC O01149;
DT 01-APR-1993 (Rel. 25, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 15-MAR-2004 (Rel. 43, Last annotation update)
DE Collagen alpha 2(I) chain precursor.
GN COL1A2 OR COLA2.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Calvaria;
RX MEDLINE=92372043; PubMed=1505972;
RA Phillips C.L., Morgan A.L., Lever L.W., Wenstrup R.J.;
RT "Sequence analysis of a full-length cDNA for the murine pro alpha
RT 2(I) collagen chain: comparison of the derived primary structure with
RT human pro alpha 2(I) collagen.";
RL Genomics 13:1345-1346(1992).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=Mammary gland;
RX MEDLINE=92372043; PubMed=12477932;
RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G., Schuler G.D.,
RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.W., Schaefer C.F., Bhat N.K.,
RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
RA Brownstein M.J., Usdin T.B., Toshiyuki S., Carninci P., Prange C.,
RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullany S.J.,
RA Bosak S.A., McSwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
RA Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
RA Fahy J., Helton E., Kettman M., Madan A., Rodriguez S., Sanchez A.,
RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
RA Blakesley R.W., Touchman J.W., Green B.D., Dickson M.C.,
RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,
RA Butterfield Y.S.N., Krzywinski M.I., Skalska U., Smallos D.E.,
RA Schnerch A., Schein J.E., Jones S.J.M., Marra M.A.;
RT "Generation and initial analysis of more than 15,000 full-length
RT human and mouse cDNA sequences.";
RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
RN [3]
RP SEQUENCE OF 1-110 FROM N.A.
RC TISSUE=Calvaria;
RX MEDLINE=92084969; PubMed=1748823;
RA Phillips C.L., Lever L.W., Pinnell S.R., Quarles L.D.,
RA Wenstrup R.J.;
RT "Construction of a full-length murine pro alpha 2(I) collagen cDNA by
RT the polymerase chain reaction.";
RL J. Invest. Dermatol. 97:980-984(1991).
RN [4]
RP SEQUENCE OF 1-23 FROM N.A.
RX MEDLINE=87289650; PubMed=3039494;
RA Rossi P., de Crombrughe B.;
RT "Identification of a cell-specific transcriptional enhancer in the
RT first intron of the mouse alpha 2 (type I) collagen gene.";
RL Proc. Natl. Acad. Sci. U.S.A. 84:5590-5594(1987).
CC -1- FUNCTION: Type I collagen is a member of group I collagen
CC (fibrillar forming collagen).
CC -1- SUBUNIT: Trimers of one alpha 2(I) and two alpha 1(I) chains.
CC -1- TISSUE SPECIFICITY: Forms the fibrils of tendon, ligaments and
CC bones. In bones the fibrils are mineralized with calcium
CC hydroxyapatites.
CC -1- PTM: Prolines at the third position of the tripeptide repeating
CC unit (G-X-Y) are hydroxylated in some or all of the chains.

```


RT 1(III)CB9B (positions 928-1028).";
RL Hoppe-Seyler's Z. Physiol. Chem. 360:861-868(1979).
CC -!- FUNCTION: Collagen type III occurs in most soft connective tissues
CC along with type I collagen.
CC -!- SUBUNIT: Trimers of identical alpha 1(III) chains. The chains are
CC linked to each other by interchain disulfide bonds. Trimers are
CC also cross-linked via hydroxylysines.
CC -!- PTM: Prolines at the third position of the tripeptide repeating
CC unit (G-X-Y) are hydroxylated in some or all of the chains.
DR PIR; A02862; CGB075.
DR InterPro; IPR008161; Clq helix.
DR InterPro; IPR008160; Collagen.
DR InterPro; IPR001007; VWF_C.
DR Pfam; PF01391; Collagen; 17.
DR ProDom; PD000007; Clq helix; 3.
DR PROSITE; PS01208; VWF_C_1; PARTIAL.
KW Extracellular matrix; Connective tissue; Repeat; Hydroxylation;
KW Glycoprotein; Collagen.
FT DOMAIN 1 14 NONHELICAL REGION (N-TERMINAL).
FT DOMAIN 15 1040 TRIPLE-HELICAL REGION.
FT DOMAIN 1041 1049 NONHELICAL REGION (C-TERMINAL).
FT MOD RES 95 95 HYDROXYLATION.
FT MOD RES 107 107 HYDROXYLATION.
FT MOD RES 119 119 HYDROXYLATION.
FT MOD RES 938 938 HYDROXYLATION.
FT MOD RES 950 950 HYDROXYLATION.
FT MOD RES 107 107 O-LINKED (GAL. . .).
FT CARBOHYD 950 950 O-LINKED (GAL. . .).
FT DISULFID 1040 1040 INTERCHAIN.
FT DISULFID 1041 1041 INTERCHAIN.
SQ SEQUENCE 1049 AA; 93651 MW; 8EEC33D1C66EC9A3 CRC64;
Query Match 62.1%; Score 1905.5; DB 1; Length 1049;
Best Local Similarity 62.7%; Pred. No. 9.1e-71;
Matches 347; Conservative 36; Mismatches 161; Indels 9; Gaps 1;
QY 1 GSEGEVGRGPPGPPAGAGPAGDPADEPGAKGADGAPGPPGARGPSGPE 60
DB 204 GSSGAPGQRPQGHAGAPGPPGSDGSPGKGEWGPAGIPGAPGLIGARGPPGP 263
QY 61 GPGPPGPKGDSGPPGAPGSGKDTAKGPPGPPGPPGPPGPPGPPGPPGPPG 120
DB 264 GTNGVPGQGAAGPPGKNGKAGDPGPPGPPGPPGPPGPPGPPGPPGPPG 323
QY 121 GPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPG 180
DB 324 GAAGGVPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPG 383
QY 181 GSPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPG 240
DB 384 GSPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPG 443
QY 241 GPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPG 300
DB 444 GPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPG 503
QY 301 GVPGLGAPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPG 360
DB 504 GPKGAGAPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPG 563
QY 361 GPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPG 420
DB 564 GTPGLQGMPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPG 623
QY 421 GESGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPG 480
DB 624 GESGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPG 683
QY 481 GPAGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPG 534
DB 684 GAAGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPG 743
QY 535 ---GPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPG 544

DB 744 GKDGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPG 756
RESULT 15
CAL3_HUMAN
ID CAL3_HUMAN STANDARD; PRT; 1466 AA.
AC F02461; Q15112;
DT 21-JUL-1986 (Rel. 01, Created)
DT 01-JAN-1990 (Rel. 13, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Collagen alpha 1(III) chain precursor.
GN COL3A1.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Skin fibroblast;
RX MEDLINE=89350838; PubMed=2764886;
RA Ala-Kokko L., Kontusaari S., Baldwin C.T., Kuivaniemi H.,
RA Prockop D.J.;
RT "Structure of cDNA clones coding for the entire prepro alpha 1 (III)
RT chain of human type III procollagen. Differences in protein structure
RT from type I procollagen and conservation of codon preferences.";
RL Biochem. J. 260:509-516(1989).
RN [2]
RP SEQUENCE OF 149-1225 FROM N.A.
RX MEDLINE=89386015; PubMed=2780304;
RA Janeczko R.A., Ramirez F.;
RT "Nucleotide and amino acid sequences of the entire human alpha 1
RT (III) collagen.";
RL Nucleic Acids Res. 17:6742-6742(1989).
RN [3]
RP SEQUENCE OF 168-398.
RX MEDLINE=77134724; PubMed=557335;
RA Seyer J.M., Kang A.H.;
RT "Covalent structure of collagen: amino acid sequence of cyanogen
RT bromide peptides from the amino-terminal segment of type III collagen
RT of human liver.";
RL Biochemistry 16:1158-1164(1977).
RN [4]
RP REVISIONS.
RA Seyer J.M.;
RN Submitted (DEC-1977) to the PIR data bank.
RN [5]
RP SEQUENCE OF 399-727.
RX MEDLINE=79000343; PubMed=687591;
RA Seyer J.M., Kang A.H.;
RT "Covalent structure of collagen: amino acid sequence of five
RT consecutive CNBr peptides from type III collagen of human liver.";
RL Biochemistry 17:3404-3411(1978).
RN [6]
RP SEQUENCE OF 728-964.
RX MEDLINE=80198282; PubMed=6246925;
RA Seyer J.M., Mainardi C., Kang A.H.;
RT "Covalent structure of collagen: amino acid sequence of alpha 1
RT (III)-CB8 from type III collagen of human liver.";
RL Biochemistry 19:1583-1589(1980).
RN [7]
RP SEQUENCE OF 950-1466 FROM N.A.
RX MEDLINE=88189827; PubMed=3357782;
RA Mankoo B.S., Dalgleish R.;
RT "Human pro alpha 1(III) collagen: cDNA sequence for the 3' end.";
RL Nucleic Acids Res. 16:2337-2337(1988).
RN [8]
RP REVISION TO 1184.
RX MEDLINE=89098346; PubMed=3211760;
RA Molyneux K., Dalgleish R.;
RT "Human type III collagen 'variant' is a cDNA cloning artefact.";
RL Nucleic Acids Res. 16:11833-11833(1988).
RN [9]

RP SEQUENCE OF 1065-1466 FROM N.A.
 RX MEDLINE=85087944; PubMed=6096827;
 RA Loidl H.R., Brinker J.M., May M., Pihlajaniemi T., Morrow S.,
 RA Rosenbloom J., Myers J.C.;
 RT "Molecular cloning and carboxyl-propeptide analysis of human type III
 RT procollagen";
 RL Nucleic Acids Res. 12:9383-9394(1984).
 RN [10]
 RP SEQUENCE OF 965-1200.
 RX MEDLINE=81208139; PubMed=7016180;
 RA Seyer J.M., Kang A.H.;
 RT "Covalent structure of collagen: amino acid sequence of alpha
 RT 1(III)-CB9 from type III collagen of human liver";
 RL Biochemistry 20:2621-2627(1981).
 RN [11]
 RP SEQUENCE OF 1176-1466 FROM N.A.
 RX MEDLINE=85157600; PubMed=2579949;
 RA Chu M.-L., Weil D., de Wet W.J., Bernard M.P., Sippola M., Ramirez F.;
 RT "Isolation of cDNA and genomic clones encoding human pro-alpha 1
 RT (III) collagen. Partial characterization of the 3' end region of the
 RT gene";
 RL J. Biol. Chem. 260:4357-4363(1985).
 RN [12]
 RP SEQUENCE OF 1161-1200 FROM N.A.
 RX MEDLINE=86187804; PubMed=3754462;
 RA Mieskulin M., Dalglish R., Kluge-Beckerman B., Rennard S.I.,
 RA Tolstoshev P., Brantly M., Crystal R.G.;
 RT "Human type III collagen gene expression is coordinately modulated
 RT with the type I collagen genes during fibroblast growth";
 RL Biochemistry 25:1408-1413(1986).
 RN [13]
 RP SEQUENCE OF 1-170 FROM N.A.
 RC TISSUE=Placenta;
 RX MEDLINE=88303360; PubMed=3405773;
 RA Toman D., Ricca G., de Crombrughe B.;
 RT "Nucleotide sequence of a cDNA coding for the amino-terminal region
 RT of human prepro alpha 1(III) collagen";
 RL Nucleic Acids Res. 16:7201-7201(1988).
 RN [14]
 RP SEQUENCE OF 1-176 FROM N.A.
 RX MEDLINE=89378752; PubMed=2777083;
 RA Benson-Chanda V., Su M.W., Weil D., Chu M.-L., Ramirez F.;
 RT "Cloning and analysis of the 5' portion of the human type-III
 RT procollagen gene (COL3A1)";
 RL Gene 78:255-265(1989).
 RN [15]
 RP REVIEW ON VARIANTS.
 RX MEDLINE=97255959; PubMed=9101290;
 RA Kuivaniemi H., Tromp G., Prockop D.J.;
 RT "Mutations in fibrillar collagens (types I, II, III, and XI), fibril-
 RT associated collagen (type IX), and network-forming collagen (type X)
 RT cause a spectrum of diseases of bone, cartilage, and blood vessels";
 RL Hum. Mutat. 9:300-315(1997).
 RN [16]
 RP VARIANT AORTIC ANEURYSM ARG-303, AND VARIANT THR-668.
 RX MEDLINE=9293998; PubMed=8514866;
 RA Tromp G., Wu Y., Prockop D.J., Madhathari S.L., Kleinert C.,
 RA Earley J.J., Zhuang J., Noerregaard O., Darling R.C., Abbott W.M.,
 RA Cole C.W., Jaakkola P., Ryyanen M., Pearce W.H., Yao J.S.T.,
 RA Majamaa K., Smullen S.V., Gatalica Z., Ferrell R.E., Jimenez S.A.,
 RA Jackson C.E., Michels V.V., Kaye M., Kuivaniemi H.;
 RT "Sequencing of cDNA from 50 unrelated patients reveals that mutations
 RT in the triple-helical domain of type III procollagen are an
 RT infrequent cause of aortic aneurysms";
 RL J. Clin. Invest. 91:2539-2545(1993).
 RN [17]
 RP VARIANT THR-698.
 RX MEDLINE=91045136; PubMed=2235526;
 RA Zafarullah K., Kleinert C., Tromp G., Kuivaniemi H., Kontusaari S.,
 RA Wu Y., Ganguly A., Prockop D.J.;
 RT "G to A polymorphism in exon 31 of the COL3A1 gene";
 RL Nucleic Acids Res. 18:6180-6180(1990).
 RN [18]

RP VARIANT AORTIC ANEURYSM ARG-786.
 RX MEDLINE=91056145; PubMed=2243125;
 RA Kontusaari S., Tromp G., Kuivaniemi H., Romanic A.M., Prockop D.J.;
 RT "A mutation in the gene for type III procollagen (COL3A1) in a family
 RT with aortic aneurysms";
 RL J. Clin. Invest. 86:1465-1473(1990).
 RN [19]
 RP VARIANT EDS-IV ARG-828.
 RX MEDLINE=94016385; PubMed=8411057;
 RA Richards A.J., Narcisi P., Lloyd J.C., Ferguson C., Pope F.M.;
 RT "The substitution of glycine 861 by arginine in type III collagen
 RT produces mutant molecules with different thermal stabilities and
 RT causes Ehlers-Danlos syndrome type IV";
 RL J. Med. Genet. 30:690-693(1993).
 RN [20]
 RP VARIANT EDS-IV SER-957.
 RX MEDLINE=89109135; PubMed=2492273;
 RA Tromp G., Kuivaniemi H., Shikata H., Prockop D.J.;
 RT "A single base mutation that substitutes serine for glycine 790 of
 RT the alpha 1 (III) chain of type III procollagen exposes an arginine
 RT and causes Ehlers-Danlos syndrome IV";
 RL J. Biol. Chem. 264:1349-1352(1989).
 RN [21]
 RP VARIANT EDS-IV VAL-960.
 RX MEDLINE=95268429; PubMed=7749417;
 RA Tromp G., de Paeppe A., Nuytink L., Madhathari S.L., Kuivaniemi H.;
 RT "Substitution of valine for glycine 793 in type III procollagen in
 RT Ehlers-Danlos syndrome type IV";
 RL Hum. Mutat. 5:179-181(1995).
 RN [22]
 RP VARIANT EDS-IV GLU-1014.
 RX MEDLINE=92316511; PubMed=1352273;
 RA Richards A.J., Ward P.N., Narcisi P., Nicholls A.C., Lloyd J.C.,
 RA Pope F.M.;
 RT "A single base mutation in the gene for type III collagen (COL3A1)
 RT converts glycine 847 to glutamic acid in a family with Ehlers-Danlos
 RT syndrome type IV. An unaffected family member is mosaic for the
 RT mutation";
 RL Hum. Genet. 89:414-418(1992).
 RN [23]
 RP VARIANT EDS-IV ASP-1050.
 RX MEDLINE=90037070; PubMed=2808425;
 RA Tromp G., Kuivaniemi H., Stolle C.A., Pope F.M., Prockop D.J.;
 RT "Single base mutation in the type III procollagen gene that converts
 RT the codon for glycine 883 to aspartate in a mild variant of
 RT Ehlers-Danlos syndrome IV";
 RL J. Biol. Chem. 264:19313-19317(1989).
 RN [24]
 RP VARIANT EDS-IV VAL-1077.
 RX MEDLINE=91374480; PubMed=1895316;
 RA Richards A.J., Lloyd J.C., Ward P.N., de Paeppe A., Narcisi P.,
 RA Pope F.M.;
 RT "Characterisation of a glycine to valine substitution at amino acid
 RT position 910 of the triple helical region of type III collagen in a
 RT patient with Ehlers-Danlos syndrome type IV";
 RL J. Med. Genet. 28:458-463(1991).
 RN [25]
 RP MEDLINE=93022543; PubMed=1357232;
 RA Johnson P.H., Richards A.J., Pope F.M., Hopkinson D.A.;
 RT Query Match 61.9%; Score 1901; DB 1; Length 1466;
 RT Best Local Similarity 63.1%; Pred. No. 1,7e-70;
 RT Matches 345; Conservative 37; Mismatches 159; Indels. 6; Gaps 2;
 QY 1 GSEGEVGRGPPGPPGAGAGGADGAGCGAKGADGAGTAGAGPFGPGRSGPE 60
 DB 360 GSNAGAPGORGPPGQGHAGACGPPGPPGNGSPGCKGEMGAPGPGAGLNGARGPPGPA 419
 QY 61 QPGPPGPKGDSGEPGAPGKGTGAKGEPGVPVGEPGPGAGEGKPGARGPGFTGLP 120
 DB 420 GAGAPGLRGAGGEPGKNGKGPGRGERGAGIPGVPKAGKGDGKDSGSPGPNGLP 479

[illegible]

Search completed: September 24, 2004, 11:07:24
Job time : 14.7264 secs

This Page Blank (uspto)

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: September 24, 2004, 11:06:55 ; Search time 74.4117 Seconds
(without alignments)
2065.614 Million cell updates/sec

Title: US-10-658-989A-4
Perfect score: 3070
Sequence: 1 GSECPGVRGEPGPAG.....PGPSGDAGPPGPPGAGKEG 544

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 1586107 seqs, 282547505 residues

Total number of hits satisfying chosen parameters: 1586107

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : A_Geneseq_29Jan04.*
1: Geneseqp1980s.*
2: Geneseqp1990s.*
3: Geneseqp2000s.*
4: Geneseqp2001s.*
5: Geneseqp2002s.*
6: Geneseqp2003as.*
7: Geneseqp2003bs.*
8: Geneseqp2004s.*

ALIGNMENTS

RESULT 1
AAY84541
ID AAY84541 standard; protein; 1057 AA.
XX AC AAY84541;
XX DT 25-JUL-2000 (first entry)
XX DE Amino acid sequence of a human collagen 1 (alpha1) protein.
XX KW Extracellular matrix protein; self aggregation; hydroxylated proline;
XX KW trans-4-hydroxyproline; 3-hydroxyproline; recombinant protein production;
XX KW collagen; fibrinogen; fibronectin; post translational hydroxylation.
XX OS Homo sapiens.
XX PN EP992586-A2.
XX PD 12-APR-2000.
XX PF 07-OCT-1999; 99EP-00119184.
XX PR 09-OCT-1998; 98US-00169769.
(USSU) US SURGICAL CORP.
Gruskin EA, Buechter DD, Zhang G, Connolly K;
WPI; 2000-259138/23.
N-PSDB; AAA12502.
Production of extracellular matrix proteins containing 4-trans-
hydroxyproline results in native self aggregating proteins, useful on
medical implants.
Disclosure; Fig 27A-E; 260pp; English.
The specification describes a method for producing an extracellular
matrix protein or its fragment. The extracellular matrix protein is
capable of self aggregating in a cell which does not ordinarily
hydroxylated prolines. The method comprises optimising a nucleic acid
sequence for expression in the cell by substitution of codons preferred
by that cell for naturally occurring codons not preferred by the cell;
incorporating the nucleic acid sequence into the cell; and contacting the
cell with a hypertonic growth medium containing at least one amino acid,
selected from the group consisting of trans-4-hydroxyproline and 3-
hydroxyproline to allow at least one of the amino acids to be assimilated
into the cell and incorporated into the extracellular matrix protein. The

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	2966	96.6	1057	3 AAY84541	AAY84541 Amino aci
2	2966	96.6	1057	3 AAY84544	AAY84544 A human c
3	2966	96.6	1058	3 AAY84403	AAY84403 Amino aci
4	2966	96.6	1107	2 AAR8472	AAR8472 Collagen/
5	2966	96.6	1107	3 AAY84540	AAY84540 Amino aci
6	2966	96.6	1161	7 ADE87050	ADE87050 Human pan
7	2966	96.6	1169	2 AAR89469	AAR89469 Collagen/
8	2966	96.6	1169	3 AAY84537	AAY84537 Amino aci
9	2966	96.6	1171	2 AAR89470	AAR89470 Collagen/
10	2966	96.6	1171	3 AAY84538	AAY84538 A chimeri
11	2966	96.6	1211	7 ADE87057	ADE87057 Human pan
12	2966	96.6	1226	7 ADE87062	ADE87062 Human pan
13	2966	96.6	1388	2 AAR89471	AAR89471 Collagen/
14	2966	96.6	1411	3 AAY56800	AAY56800 Human pre
15	2966	96.6	1461	5 ABG93947	ABG93947 Human pol
16	2966	96.6	1464	2 AAW68485	AAW68485 Human rec
17	2966	96.6	1464	4 AAB82454	AAB82454 Human pro
18	2966	96.6	1464	4 AAU14136	AAU14136 Human nov
19	2966	96.6	1464	5 ABB90764	ABB90764 Human tum
20	2966	96.6	1464	5 ABP68610	ABP68610 Human pan
21	2966	96.6	1464	6 ABU54471	ABU54471 Human tum
22	2966	96.6	1464	6 ABR47417	ABR47417 Breast ca
23	2966	96.6	1464	6 ABR92064	ABR92064 Human cer
24	2966	96.6	1464	7 ADD14142	ADD14142 Human src
25	2966	96.6	1464	7 ADD45059	ADD45059 Human Pro

Best Local Similarity		95.6%	Pred. No. 2.8e-169;	
Matches 520; Conservative		21;	Mismatches 3;	Indels 0; Gaps 0;

QY	1	GSEGEVGRPEGPPGPPAGAAAGPAGDPCADGEPKAGKADGAPGIAGAPFPFGAGPSGPE	60
Db	201	GSESPQGVGEFPGPPGAGAAAGPAGNFGADGQPGKAGKANGAPGIAGAPFPFGAGPSGPQ	260
QY	61	GPQGGPPKGDSEFPGAPSGKDTGAKGEPQPVGVEPPGPAGBEGKPGARGEPGPTGLP	120
Db	261	GPQGGPPKGNSEFPGAPSGKDTGAKGEPQFVGQPPGPGABEGKRGARGEPGPTGLP	320
QY	121	GPPEERGPPSGRSGFPFGADGVAGPKGPPAGERGSPGAPFKGSPGAGRPGEAGLPCAAGLT	180
Db	321	GPPEERGPPSGRSGFPFGADGVAGPKGPPAGERGSPGAPFKGSPGAGRPGEAGLPCAAGLT	380
QY	181	GSPGSPGPDCKTGPPGPAAGEDRGPPGPPGPPGARGEAQVMGPPGPKGAAGEPKKAGRGVP	240
Db	381	GSPGSPGPDCKTGPPGPAAGEDRGPPGPPGPPGARGEAQVMGPPGPKGAAGEPKKAGRGVP	440
QY	241	GPPGAVGPAGKDGESAGAEAGPPGPPGAPGAGERGEEGPPGSGPEGLPGPAGPPGEAGKPGEE	300
Db	441	GPPGAVGPAGKDGESAGAEAGPPGPPGAPGAGERGEEGPPGSGPEGLPGPAGPPGEAGKPGEE	500
QY	301	GVPGDLGAPGSPGARGPPGPPGGERGVGPPGPPGAPGPGADGAPGDDGAKGADAGAPGAPGSE	360
Db	501	GVPGDLGAPGSPGARGPPGPPGGERGVGPPGPPGAPGPGADGAPGDDGAKGADAGAPGAPGSE	560
QY	361	GAPLEGMPGERGAAGLPGPKRGDAGPKGADGSPGKGVRGHTGPIGPPGPPGAPGDX	420
Db	561	GAPLEGMPGERGAAGLPGPKRGDAGPKGADGSPGKGVRGHTGPIGPPGPPGAPGDX	620
QY	421	GESGSPGFAGPTGARGAPGDRGEPGPPGPPGAGFAGPPGADGEPGAKGEPGPDAGAKGDAGPP	480
Db	621	GESGSPGFAGPTGARGAPGDRGEPGPPGPPGAGFAGPPGADGEPGAKGEPGPDAGAKGDAGPP	680
QY	481	GPAGPAGPPGPIGDVAGAPKAGKAGSGAGPPGATGFPGAAGRVGPPGSGDAGPPGPPGPA	540
Db	681	GPAGPAGPPGPIGDVAGAPKAGKAGSGAGPPGATGFPGAAGRVGPPGSGDAGPPGPPGPA	740
QY	541	GKSG 544	
Db	741	GKSG 744	

RESULT 6		
AD87050	AD87050	
ID	AD87050 standard; protein; 1161 AA.	
AC	XX	
AC	XX	
XX	AD87050;	
XX		
DT	29-JAN-2004 (first entry)	
DE	Human pancreatic cell protein sequence SegID510.	
DE	neoplastic pancreatic cell; pancreatic cell; pancreatic cancer;	
KW	cancer death; cytostatic; vaccine; gene therapy;	
KW	non-cancerous pancreas disease; human.	
XX		
OS	Homo sapiens.	
XX		
XX	WO2003060145-A2.	
PD	24-JUL-2003.	
XX		
PF	19-DEC-2002; 2002WO-US040655.	
XX		
PR	21-DEC-2001; 2001US-0342768P.	
XX		
PA	(DIAD-) DIADEXUS INC.	
XX		
PI	Sun Y, Liu C;	
XX	WFI; 2003-587286/55.	

DR N-PSDB: AD887387.
XX
XX New pancreatic specific nucleic acid molecule or protein for diagnosing,
PT staging, imaging, monitoring, preventing or treating pancreatic cancer or
PT non-cancerous disease states of the pancreas.
XX
XX
PS Claim 12; SEQ ID NO 510; 635bp; English.
XX
XX This invention relates to novel nucleic acids and proteins present in
CC normal and neoplastic pancreatic cells. Pancreatic cancer is a common
CC cause of cancer death worldwide, therefore accurate methods of diagnosis
CC and treatment are required. Compounds which modulate the proteins of the
CC invention may have cytostatic activity and the protein and DNA sequences
CC of the invention may be useful for the development of a vaccine or in
CC gene therapy. The composition and methods are useful in diagnosing,
CC staging, imaging, monitoring, preventing or treating pancreatic cancer
CC and non-cancerous disease states of the pancreas. The present sequence is
CC that of a human pancreatic protein of the invention.

SQ Sequence 1161 AA; Query Match 96.6%; Score 2966; DB 7; Length 1161;
Best Local Similarity 95.6%; Pred. No. 2.9e-169;
Matches 520; Conservative 21; Mismatches 3; Indels 0; Gaps 0

Qy	1	GSEGEVGRGPPGPPGAGAGPADPGADGEPGAKGADGAPGIAAGAPFFPGAGPSGE	60
Db	59	GSEGPQGVGRGPPGPPGAGAGPAGNPGADQCPGAKGANGAPGAGAPFPGARGSPGQ	118
Qy	61	GSGPPGPKGDSGPPGAPSGKDDTGAKGEGPVGVPGPAPAGERKPGARGEPGTGLP	120
Db	119	GPGGPPGPKGNSGPPGAPSGKDDTGAKGEGPVGVPGPAPAGEBKRGARGEPGTGLP	178
Qy	121	GPPGERGGPGRGPPGADGVAGPKGAPAGERGSPGAPKSGPGEAGRPGEAGLPGAKGLT	180
Db	179	GPPGERGGPGRGPPGADGVAGPKGAPAGERGSPGAPKSGPGEAGRPGEAGLPGAKGLT	238
Qy	181	GSPGSGPDGKTGPPGPAGEDRGPPGAPGAGEAGVMGFPQPKGAAGEPGKAGRGVP	240
Db	239	GSPGSGPDGKTGPPGPAGQDGRPPGPPPCARQAGVMGFPQPKGAAGEPGKAGRGVP	298
Qy	241	GPPGAVGAGKDXGAGABGPPGAPAGERGEGSPAGSGPCEGLPGPAGPPGEGAGKPGEE	300
Db	299	GPPGAVGAGKDXGAGABGPPGAPAGERGEGSPAGSGPCEGLPGPAGPPGEGAGKPGEE	358
Qy	301	GVFGDLGAPGSGARGGPPGGERGVSGPPGAPGPPGADGAPDGDGAKGDGAPGAPGSE	360
Db	359	GVFGDLGAPGSGARGGPPGGERGVSGPPGAPGPPGADGAPDGDGAKGDGAPGAPGSO	418
Qy	361	GAPLUEGVPGBRGAGLPGPKDGRGDGAPKXAGDSGPKXGQVGRGLTGPIGPPGAPAGDX	420
Db	419	GAPLQGGVPGERGAGLPKPKDGRGDGAPKXAGDSGPKXGQVGRGLTGPIGPPGAPAGDX	478
Qy	421	GESGSGPAGTARGAPGDRGEPGPPGAPGAPGADGEPGAKGCEGCDGAKGDGADPP	480
Db	479	GESGSGPAGTARGAPGDRGEPGPPGAPGAPGADGEPGAKGCEGCDGAKGDGADPP	538
Qy	481	GFPAGPAGPPGIGDVGAFGAKGARGSGPPGATGFPFGAAGRVGPPGSGDAGPPGPPGA	540
Db	539	GFPAGPAGPPGIGDVGAFGAKGARGSGPPGATGFPFGAAGRVGPPGSGDAGPPGPPGA	598
Qy	541	GKEG 544	
Db	599	GKEG 602	

RESULT 7
AAR89469
ID AAR89469 standard; protein; 1189 AA.
XX
XX AAR89469;
XX
DT 01-OCT-1996 (first entry)

The specification describes a method for producing an extracellular matrix protein or its fragment. The extracellular matrix protein is capable of self aggregating in a cell which does not ordinarily hydroxylate prolines. The method comprises optimising a nucleic acid sequence for expression in the cell by substitution of codons preferred by that cell for naturally occurring codons not preferred by the cell; incorporating the nucleic acid sequence into the cell; and contacting the cell with a hypertonic growth medium containing at least one amino acid, selected from the group consisting of trans-4-hydroxyproline and 3-hydroxyproline to allow at least one of the amino acids to be assimilated into the cell and incorporated into the extracellular matrix protein. The method may be used to make host cells assimilate and incorporate trans-4-hydroxyproline into proteins. This is especially useful in the recombinant production of proteins such as collagen, fibrinogen and fibronectin whose ability to self aggregate and produce functional proteins depends on the post translational hydroxylation of proline. The method is also useful in studying the structure and function of polypeptides which do not normally contain trans-4-hydroxyproline. The present sequence represents a chimeric collagen 1 (alpha1) bone morphogenic protein-2B (bmp-2b) protein, which may be produced using the method of the invention

Sequence 1169 AA:

Query Match	96.6%;	Score 2966;	DB 3;	Length 1169;
Best Local Similarity	95.6%;	Pred. No. 2.9e-169;		
Matches	520;	Conservative 21;	Mismatches 3;	Indels 0; Gaps 0;

QY	1	GSEGEVGRGEP	PPG	PAG	AGADG	AGDGE	PCAK	GADG	AGGTAG	AGP	PGP	AG	PGSG	PE	60	
DB	201	GSQGFQVRGEP	PPG	PAG	AGAGP	AGN	PCADG	QPKA	GANGAG	PGTAG	APFP	GARG	PGSG	GP	260	
QY	61	GPQGPFPFKD	SE	PGAP	SGSKD	TG	KARGE	PGPV	GVEG	PPG	PAG	BE	GP	KARG	EGP	120
DB	261	PGQGPFPKNS	GE	PGAP	SGSKD	TG	KARGE	PGPV	QVGP	PPG	PAG	BE	GP	KARG	EGP	320
QY	121	GPQERGGP	SGSR	FP	PGAD	GV	AGPK	GP	AGER	SG	PG	AG	RG	SG	PKGL	180
DB	321	GPQERGGP	SGSR	FP	PGAD	GV	AGPK	GP	AGER	SG	PG	AG	RG	SG	PKGL	380
QY	181	GPSGSPG	CDGK	TG	PPG	AGED	RG	PPG	PPG	ARG	EA	GV	MG	FP	PG	240
DB	381	GPSGSPG	CDGK	TG	PPG	AGED	RG	PPG	PPG	ARG	EA	GV	MG	FP	PG	440
QY	241	GPPGAV	GP	AKD	GE	AG	AE	GG	PPG	PAG	AGER	GE	GP	AG	SG	300
DB	441	GPPGAV	GP	AKD	GE	AG	AE	GG	PPG	PAG	AGER	GE	GP	AG	SG	500
QY	301	GVPGD	L	CA	PPS	G	ARG	EP	PP	PC	ERG	V	GP	PG	AG	360
DB	501	GVPGD	L	CA	PPS	G	ARG	EP	PP	PC	ERG	V	GP	PG	AG	560
QY	361	GAPLEG	MP	GER	GA	AG	LP	GK	DR	D	AG	PK	G	AD	SG	420
DB	561	GAPLEG	MP	GER	GA	AG	LP	GK	DR	D	AG	PK	G	AD	SG	620
QY	421	GESG	SP	AG	PT	G	ARG	A	GP	R	GE	PP	GP	AG	F	480
DB	621	GESG	SP	AG	PT	G	ARG	A	GP	R	GE	PP	GP	AG	F	680
QY	481	GPAG	PP	GP	PI	D	VG	AP	G	AK	G	AG	SG	AG	PC	540
DB	681	GPAG	PP	GP	PI	D	VG	AP	G	AK	G	AG	SG	AG	PC	740
QY	541	GKEG	544													
DB	741	GKEG	744													

RESULT 9	AA889470
ID	AA889470
standard;	protein: 1171 AA.

XX	AAR89470;
AC	
XX	01-OCT-1996 (first entry)
DT	
XX	
DE	Collagen/TGF-beta-1 fusion protein.
XX	
KW	Transforming growth factor; TGF-beta-1; collagen IA; osteogenesis;
KX	bore formation; tissue repair; fusion protein.
XX	
OS	Synthetic.
XX	
Key	Location/Qualifiers
PH	1..1057
FT	/label= Collagen-IA
FT	/note= "collagen IA alpha-helical domain"
FT	Misc-difference 887
FT	/note= "unidentified amino acid"
FT	Misc-difference 890
FT	/note= "unidentified amino acid"
FT	1058..1059
FT	/label= Linker_peptide
FT	1060..1171
FT	/label= TGF-beta-1
FT	/note= "human mature TGF-beta-1"
XX	
PN	CA2151547-A..
XX	
PD	11-DEC-1995..
XX	
PP	12-JUN-1995; 95CA-02151547.
XX	
PR	10-JUN-1994; 94US-00259263.
XX	
PA	(USSU) US SURGICAL CORP.
XX	
PI	Gruskin EA, Espino P;
XX	
DR	WPI; 1996-140144/15.
DR	N-PSDB: AAT16516.

XX	Chimeric DNA encoding protein contg. extracellular matrix protein domain
PT	- and cellular regulatory factor domain, partic. useful as osteogenic
PT	agents, also related vectors, transformed cells and chimaeric proteins.
XX	
PS	Disclosure; Fig 6; 59pp; English.
XX	
CC	A fusion protein (AAR89470) comprises the alpha-helical region of human
CC	collagen I(a) linked to the human mature transforming growth factor beta
CC	I (TGF-beta-1). It can be expressed in Escherichia coli transformants
CC	carrying a vector incorporating a chimeric gene (AA161516) coding for the
CC	fusion. The TGF-beta-moiety increases efficacy of the body's normal soft
CC	tissue repair response and also induces osteogenesis. The collagen moiety
CC	provides an integral substratum or scaffolding for the TGF and cells
CC	involved in reconstruction and growth. The fusion protein provides
CC	sustained release and delivery of TGF-beta-1 to a target tissue
XX	
SQ	Sequence 1171 AA;
	Query Match 96.6%; Score 2966; DB 2; Length 1171;
	Best Local Similarity 95.6%; Pred.No. 2.9e-169;
	Matches 520; Conservative 21; Mismatches 0; Gaps 0; Gaps 0;
QY	1 GSEGGVGRGPPGPPAGAAAGDPDGDGECAKAGDAPCIACAGFPFGARGSPGPE 60
	: :
DB	201 GSEGGVGRGPPGPPGAGAGPAGNFDGADQFCAKANGAPCIACAGFPFGARGSPGPQ 260
	: :
QY	61 GGCGPPGPKGDSGPGAPSGKDTGAKEGPPGVGVGGPPCPAGEEKGPCAREPPTGLP 120
	: ~:
DB	261 GGCGPPGPKGNSGPFAPSGSKDGTGAKGEPGVGVGGPPGPAEECKRGAREGPPGTGLP 320
	: ~:
QY	121 GPFFRERGPGSRGPPGADVAGPKPGAGERGSGPPAGPKSGPEACRGPBAGLPCAAGILT 180

ID ADE87057 standard; protein; 1211 AA.
AC ADE87057;
XX
DT 29-JAN-2004 (first entry)
XX
XX Human pancreatic cell protein sequence SeqID517.
DE
XX neoplastic pancreatic cell; pancreatic cell; pancreatic cancer;
KW cancer death; cytostatic; vaccine; gene therapy;
XX non-cancerous pancreas disease; human.
XX
OS Homo sapiens.
XX
XX WO2003060145-A2.
PN
XX 24-JUL-2003.
PD
XX 19-DEC-2002; 2002WO-US040655.
PF
XX 21-DEC-2001; 2001US-0342768P.
PR
XX (DIAD-) DIADEXUS INC.
XX
XX Sun Y, Liu C;
PI
XX WPI: 2003-587286/55.
DR N-PSDB; ADE87397.
DR
XX New pancreatic specific nucleic acid molecule or protein for diagnosing,
PT staging, imaging, monitoring, preventing or treating pancreatic cancer or
PT non-cancerous disease states of the pancreas.
XX
PS Claim 12; SEQ ID NO 517; 635pp; English.
XX
XX This invention relates to novel nucleic acids and proteins present in
CC normal and neoplastic pancreatic cells. Pancreatic cancer is a common
CC cause of cancer death worldwide, therefore accurate methods of diagnosis
CC and treatment are required. Compounds which modulate the proteins of the
CC invention may have cytostatic activity and the protein and DNA sequences
CC of the invention may be useful for the development of a vaccine or in
CC gene therapy. The composition and methods are useful in diagnosing,
CC staging, imaging, monitoring, preventing or treating pancreatic cancer
CC and non-cancerous disease states of the pancreas. The present sequence is
CC that of a human pancreatic protein of the invention.
XX
SQ Sequence 1211 AA;

Query Match 96.6%; Score 2966; DB 7; Length 1211;
Best Local Similarity 95.6%; Pred. No. 3e-169;
Matches 520; Conservative 21; Mismatches 3; Indels 0; Gaps 0;

QY 1 GSEGEVGRGEPGPPGAGAGPAGDPCADCEFGAKGADGAPGAPGPPGARGSPGPE 60
DB 362 GSEGPQGVGRGEPGPPGAGAGPAGNPGADGQFGKANGAPGAGPPGARGSPGQ 421

QY 61 GPGGPPGKDSBEPGAPGSKGDTGAKGEPGVPVGVGPPGAPGEGKPGARGPPTGLP 120
DB 422 GPGGPPGKNSBEPGAPGSKGDTGAKGEPGVPVGVGPPGAPGEGKPGARGPPTGLP 481

QY 121 GPPGEGGSGSRGPPGADGVAGPKGAPGARGSGPPGAPGKSGPCEAGRPGAGLPGAKGLT 180
DB 482 GPPGEGGSGSRGPPGADGVAGPKGAPGARGSGPPGAPGKSGPCEAGRPGAGLPGAKGLT 541

QY 181 GSPGSPGDKTGPFGAGDGRPPGPPGARGGAGVMGPPGKGAAGEPPGKAGRGVP 240
DB 542 GSPGSPGDKTGPFGAGDGRPPGPPGARGGAGVMGPPGKGAAGEPPGKAGRGVP 601

QY 241 GPPGAVGPKGDSBEPGAPGARGSGPCEAGRPGAGLPGAPGPPGARGPPEE 300
DB 602 GPPGAVGPKGDSBEPGAPGARGSGPCEAGRPGAGLPGAPGPPGARGPPEE 661

QY 301 GVPDGLGAPGSGARGPFGPPGARGVEGPPGAPGADGADGAKGDAGAPGPGSE 360

DB 662 GVPDGLGAPGSGARGPFGPPGARGVEGPPGAPGADGADGAKGDAGAPGPGSQ 721
QY 361 GAPGLEGNPFGERRGAAGLPKPKGDRGADGPKGADGSPKDGVRGLTGPIGPPGAPGADK 420
DB 722 GAPGLQGNPFGERRGAAGLPKPKGDRGADGPKGADGSPKDGVRGLTGPIGPPGAPGADK 781
QY 421 GSEFSGPAGTGAAGAPGDRGEPGPPGAPGAGPAGDGPAGKBPAGDAGADGPP 480
DB 782 GSEFSGPAGTGAAGAPGDRGEPGPPGAPGAGPAGDGPAGKBPAGDAGADGPP 841
QY 481 GPAGPAGPPGPIGVGAPGAKGARGSPGATGPPGAAGRVGPPGSGDAGPPGPPGPA 540
DB 842 GPAGPAGPPGPIGVGAPGAKGARGSPGATGPPGAAGRVGPPGSGDAGPPGPPGPA 901
QY 541 GKEG 544
DB 902 GKEG 905

RESULT 12
ADE87062
ID ADE87062 standard; protein; 1226 AA.
XX
AC ADE87062;
XX
DT 29-JAN-2004 (first entry)
XX
DE Human pancreatic cell protein sequence SeqID522.
XX
XX neoplastic pancreatic cell; pancreatic cell; pancreatic cancer;
KW cancer death; cytostatic; vaccine; gene therapy;
XX non-cancerous pancreas disease; human.
XX
OS Homo sapiens.
XX
XX WO2003060145-A2.
PN
XX 24-JUL-2003.
PD
XX 19-DEC-2002; 2002WO-US040655.
PF
XX 21-DEC-2001; 2001US-0342768P.
PR
XX (DIAD-) DIADEXUS INC.
XX
XX Sun Y, Liu C;
PI
XX WPI: 2003-587286/55.
DR N-PSDB; ADE87403.
DR
XX New pancreatic specific nucleic acid molecule or protein for diagnosing,
PT staging, imaging, monitoring, preventing or treating pancreatic cancer or
PT non-cancerous disease states of the pancreas.
XX
PS Claim 12; SEQ ID NO 522; 635pp; English.
XX
XX This invention relates to novel nucleic acids and proteins present in
CC normal and neoplastic pancreatic cells. Pancreatic cancer is a common
CC cause of cancer death worldwide, therefore accurate methods of diagnosis
CC and treatment are required. Compounds which modulate the proteins of the
CC invention may have cytostatic activity and the protein and DNA sequences
CC of the invention may be useful for the development of a vaccine or in
CC gene therapy. The composition and methods are useful in diagnosing,
CC staging, imaging, monitoring, preventing or treating pancreatic cancer
CC and non-cancerous disease states of the pancreas. The present sequence is
CC that of a human pancreatic protein of the invention.
XX
SQ Sequence 1226 AA;

Query Match 96.6%; Score 2966; DB 7; Length 1226;
Best Local Similarity 95.6%; Pred. No. 3e-169;
Matches 520; Conservative 21; Mismatches 3; Indels 0; Gaps 0;

[illegible]

Qy	181	GSFGSPGDDKTTGPPGPAAGEDGRPGPPGPGARGAEAGVMGFPDPKGAAGBPGKAGERGVP	241
Db	539	GSFGSPGDDKTTGPPGPAQDGRPGPPGPGARGAQAGVMGFPDPKGAAGBPGKAGERGVP	539
Qy	241	GGPGAVGAGKDGAGAGAGPSPGAPGAGGEEGPGAGSPGFEGLPGPAGPPGAGKPGEE	300
Db	599	GGPGAVGAGKDGAGAGAGPSPGAPGAGGEEGPGAGSPGFEGLPGPAGPPGAGKPGEE	658
Qy	301	GVPGDLGAPGPGARGEGPFFGERGVGPPGPGAPPGADGAPDDGAKGDAGAPGAPGSE	360
Db	659	GVPGDLGAPGPGARGEGPFFGERGVGPPGPGAPPGADGAPDDGAKGDAGAPGAPGSE	718
Qy	361	GAPGLEGMWGERGAAGLPQPKXDRGDAGPKGADGSPKDGVRGLTGPIGPPGAGAPGDK	420
Db	719	GAPGLEGMWGERGAAGLPQPKXDRGDAGPKGADGSPKDGVRGLTGPIGPPGAGAPGDK	778
Qy	421	GESGSPGAPPTGARGAPCDRGEPPGPPGAPGAGPPGADGEPGAKGEPGDAGAKGDAGPP	480
Db	779	GESGSPGAPPTGARGAPCDRGEPPGPPGAPGAGPPGADGEPGAKGEPGDAGAKGDAGPP	838
Qy	481	GPAGPAGPPGPDYVGAAGKARGSAGPPGATGFGAAGRVGPPGPPSGDAGPPGPPGPA	540
Db	839	GPAGPAGPPGPDYVGAAGKARGSAGPPGATGFGAAGRVGPPGPPSGDAGPPGPPGPA	898
Qy	541	GKEG 544	
Db	899	GKEG 902	
RESULT 15			
ABG93947			
ID	ABG93947	standard; protein; 1461 AA.	
AC	ABG93947;		
XX			
DT	26-NOV-2002	(first entry)	
XX			
DE		Human polypeptide orthologous to DACC-11.	
XX			
KW		Human; Deer; rat; mouse; DACC; deer antler cartilage cell;	
KW		cell stimulation; cell inhibition; cell growth; cell division;	
KW		mesenchymal cell; chondrocyte; chondrogenesis; osteogenesis; growth;	
KW		repair; regeneration; restoration; extracellular matrix;	
KW		cartilaginous matrix; cartilage; disc; connective tissue; agonist;	
XX		antagonist; gene therapy.	
OS		Homo sapiens.	
XX			
PN		W0200264625-A1.	
XX			
PD		22-AUG-2002.	
XX			
PF		15-FEB-2002; 2002WC-AU000163.	
XX			
PR		15-FEB-2001; 2001AU-00003116.	
XX			
FA		(ADPP-) ADP PHARM PTY LTD.	
PA		(UNSY) UNIV SYDNEY.	
XX			
PI		Roubin R, Ghosh P;	
XX			
DR		WPI; 2002-643456/69.	
XX			
PT		Stimulating or inhibiting cell growth and/or division, useful for	
PT		stimulating chondrogenesis, cartilage, disc or connective tissue growth,	
PT		repair, and/or regeneration, comprises administering deer antler	
PT		cartilage gene.	
XX			
PS		Claim 13; Page 154-160; 214pp; English.	
XX			
CC		The invention discloses a method for stimulating or inhibiting cell	
CC		growth and/or division which comprises contacting or inserting into an	
CC		animal cell a polypeptide comprising one of the deer antler cartilage	

RESULT 14	
AAAY56800	
ID	AAAY56800 standard; protein; 1411 AA.
XX	
XX	AAAY56800;
XX	
DT	27-MAR-2000 (first entry)
XX	
XX	Human preproalpha 1 (I) collagen.
XX	
KW	Fibrillar collagen; C propeptide; SSAD; telopeptide; gelatin;
KW	sequence selection and alignment domain; prosthetic implant; foodstuff;
KW	medicines; type I collagen; human.
XX	
OS	Homo sapiens.
XX	
FN	EP967226-A2.
XX	
PD	29-DEC-1999.
XX	
XX	04-MAY-1999; 99EP-00303470.
XX	
PR	08-MAY-1998; 98US-0084828P.
PR	10-APR-1999; 99US-00289578.
XX	
PA	(COHE-) COHESION TECHNOLOGIES INC.
XX	
PI	Olsen DR, Hitzeman RA, Chisholm GE;
XX	
XX	WPI; 2000-074666/07.
XX	
PT	New method for production of fibrillar collagen, useful for preparing
PT	telopeptide collagen fibrils and gelatin.
XX	
PS	Example 1; Fig 3A-B; 30pp; English.
XX	
CC	The invention provides a method for the production of fibrillar collagen.
CC	The method comprises: (a) culturing a recombinant host cell comprising a
CC	DNA encoding a fibrillar collagen monomer lacking a C propeptide SSAD
CC	(sequence selection and alignment domain); and (b) producing the
CC	fibrillar collagen. The methods are used to produce fibrillar collagens,
CC	from which telopeptide collagen fibrils can be derived. Host cells,
CC	comprising DNA encoding a collagen monomer lacking SSAD or N propeptide
CC	is used to produce gelatin. Collagen is used in biological research as a
CC	substrate for in vitro cell culture and as a component of biocompatible
CC	materials for use in prosthetic implants, sustained drug release
CC	matrices, artificial skin and wound dressing and healing devices. Gelatin
CC	is particularly useful for foodstuffs and medicine, for coating tablets
CC	and making capsules. The methods, comprising the use of collagen monomers
CC	lacking the N and/or C propeptides, result in a large increase in the
CC	production of type I collagen. The present sequence represents the human
CC	preproalpha 1 (I) collagen (GenBank Accn no: AF017178)
XX	
XX	Sequence 1411 AA;
XX	
Query Match	96.6%; Score 2966; DB 3; Length 1411;
Best Local Similarity	95.6%; Pred. No. 3,4e-169;
Matches 520;	Conservative 21; Mismatches 3; Indels 0; Gaps 0;
QY	1 GSGEGEVGRGEPGPPGAGAGAGADPGADGEPGAKGADGAPCIAGAPGPPGARGSPGPE 60
Db	359 GSGEGPQGVGRGEPGPPGAGAGAGAGADPGADGQPGAKGANGAPGIAGAPGPPGARGSPGQ 418
QY	61 GPGGPPGPKGDSGEPGARGSGKDTGAKGEPGVGVEGPPGAGECKPGARGEPPGTGLP 120
Db	419 GPGGPPGPKNGSGEPGARGSGKDTGAKGEPGVGVEGPPGAGECKPGARGEPPGTGLP 478
QY	121 GPPGPRGQPSRGFFGADGVGAPGKPGAGERGSPGAPGKSGPGGACRPGGAGLPKAGLT 180
Db	479 GPPGPRGQPSRGFFGADGVGAPGKPGAGERGSPGAPGKSGPGGACRPGGAGLPKAGLT 538

CC cell (DACC) clones disclosed. More particularly, the method relates to
 CC these polypeptides stimulating mesenchymal cell growth and/or division
 CC and to transfecting these cells and chondrocytes with vectors carrying
 CC the genes of these polypeptides capable of stimulating chondrogenesis,
 CC osteogenesis, growth, repair, regeneration and/or restoration of the
 CC extracellular matrix. The chondrocytes selectively express genes required
 CC to form a cartilaginous matrix. The DACC polypeptides and polynucleotides
 CC are useful for identifying an agent that modulates the activity of the
 CC polypeptide, for stimulating mesenchymal cell growth and/or division by
 CC exposing animal mesenchymal cells to conditioned media or its active
 CC fraction, obtained from deer antler cartilage cells, for inhibiting cell
 CC growth and/or division by inserting into an animal cell, a compound which
 CC inhibits the translation of the polynucleotide encoding the DACC. The
 CC method and the polypeptides are useful for stimulating mesenchymal cell
 CC growth and/or division or for stimulating chondrogenesis, cartilage, disc
 CC or connective tissue growth, repair, regeneration and/or restoration in
 CC an animal. The polynucleotides, polypeptides, agonists and antagonists
 CC may be used in treatment modalities, specifically in gene therapy. The
 CC polypeptides can be used as bait proteins in a two- or three-hybrid assay
 CC to identify other proteins, which bind to or interact with the
 CC polypeptide and are involved in modulating cell growth and/or division.
 CC The sequences presented in ABG93923-ABG93948 are the proteins encoded by
 CC the DACC cDNA clones

XX
 SQ Sequence 1461 AA;

Query Match	96.6%;	Score 2966;	DB 5;	Length 1461;
Best Local Similarity	95.6%;	Pred. No. 3.5e-169;		
Matches 520;	Conservative 21;	Mismatches 3;	Indels 0;	Gaps 0;

QY	1	GSEGEVGRGEPGPPGAGAGPAGDPGADGEPGAKGADGAPCIAGAPGPPGARGPSGE	60
DB	359	GSEGPQVGRGEPGPPGAGAGPAGDPGADGEPGAKGADGAPCIAGAPGPPGARGPSGPQ	418
QY	61	GPGGPPGPKGDSPPGAPGSGKGTGAKGEPGVPVGVGPPGAPGEGKPKGARGPPGPTGLP	120
DB	419	GPGGPPGPKGNSGPPGAPGSGKGTGAKGEPGVPVGVGPPGAPGEGKPKGARGPPGPTGLP	478
QY	121	GPPGERGGPSRGPFGADGVAGPKGAGERSGFPAGPKSGPCEAGRPCEAGLPGAKGLT	180
DB	479	GPPGERGGPSRGPFGADGVAGPKGAGERSGFPAGPKSGPCEAGRPCEAGLPGAKGLT	538
QY	181	GSPCSPGDGKTGPPGAGEDGPPGPPGARGGAGVGMGPPGKAGGPPGKAGRGVP	240
DB	539	GSPCSPGDGKTGPPGAGEDGPPGPPGARGGAGVGMGPPGKAGGPPGKAGRGVP	598
QY	241	GPPGAVGPAGKDGAGAGPPGPPGAPGAGERGEGPAGSPGEGLPFPAGPPGAGKPGEE	300
DB	599	GPPGAVGPAGKDGAGAGPPGPPGAPGAGERGEGPAGSPGEGLPFPAGPPGAGKPGEE	658
QY	301	GVPCDLGAPGSGARGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPG	360
DB	659	GVPCDLGAPGSGARGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPG	718
QY	361	GAPGLEQMPGERGAAGLPKPKGRGAGPKGADGSGPKGKGVRLTGP1GPPGPPGAPGDK	420
DB	719	GAPGLEQMPGERGAAGLPKPKGRGAGPKGADGSGPKGKGVRLTGP1GPPGPPGAPGDK	778
QY	421	GEGSPSPAGTGTARGAPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPG	480
DB	779	GEGSPSPAGTGTARGAPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPG	838
QY	481	GPAGPAGPPGPIGDVGAAPGAKGARGGAGPPGATGFFGAGRGVGPSPGSDAGPPGPPGA	540
DB	839	GPAGPAGPPGPIGDVGAAPGAKGARGGAGPPGATGFFGAGRGVGPSPGSDAGPPGPPGA	898
QY	541	GKEG 544	
DB	899	GKEG 902	

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: September 24, 2004, 11:06:56 ; Search time 22.3958 Seconds
(without alignments)
1254.011 Million cell updates/sec

Title: US-10-658-989A-4

Perfect score: 3070

Sequence: 1 GSEGGVGRGPPGPPAGA.....PGPSDAGPPGPPAGKEG 544

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 389414 seqs, 51625971 residues

Total number of hits satisfying chosen parameters: 389414

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

Issued Patents AA.*

- 1: /cgn2_6/ptodata/2/iaa/5A COMB.pdp.*
- 2: /cgn2_6/ptodata/2/iaa/5B COMB.pdp.*
- 3: /cgn2_6/ptodata/2/iaa/6A COMB.pdp.*
- 4: /cgn2_6/ptodata/2/iaa/6B COMB.pdp.*
- 5: /cgn2_6/ptodata/2/iaa/PTUS COMB.pdp.*
- 6: /cgn2_6/ptodata/2/iaa/backfiles1.pdp.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2966	96.6	1461	4	US-09-585-887-9
2	2966	96.6	1461	4	US-09-289-578-9
3	2966	96.6	1464	4	US-09-331-347C-21
4	2857	93.1	822	3	US-09-219-849-49
5	2842	92.6	1057	3	US-08-931-820-1
6	2791.5	90.9	1341	3	US-08-963-825-18
7	2791.5	90.9	1341	4	US-09-500-811-18
8	2791.5	90.9	1341	4	US-09-570-573-18
9	2791.5	90.9	1341	4	US-09-548-608-18
10	2226	72.5	1017	4	US-08-468-996-10
11	2202	71.7	1060	3	US-08-931-820-3
12	2202	71.7	1418	3	US-08-963-825-20
13	2202	71.7	1418	3	US-09-010-999-1
14	2202	71.7	1418	4	US-09-500-811-20
15	2202	71.7	1418	4	US-09-570-573-20
16	2202	71.7	1418	4	US-09-548-608-20
17	2192	71.4	1442	2	US-08-316-850-12
18	2192	71.4	1442	5	PT-US95-02251-12
19	2156	70.2	595	3	US-09-219-849-48
20	2156	70.2	595	3	US-09-219-849-50
21	1901	61.9	1057	3	US-08-931-820-4
22	1886	61.4	1024	3	US-08-931-820-2
23	1886	61.4	1365	3	US-08-963-825-19
24	1886	61.4	1365	4	US-09-500-811-19
25	1886	61.4	1366	4	US-09-570-573-19
26	1886	61.4	1366	4	US-09-548-608-19
27	1885	61.4	1366	4	US-09-585-887-10

ALIGNMENTS

RESULT 1

US-09-585-887-9

; Sequence 9, Application US/09585887

; Patent No. 6413742

; GENERAL INFORMATION:

; APPLICANT: Olsen, David R

; APPLICANT: Chang, Robert

; APPLICANT: McMullin, Hugh

; APPLICANT: Hitzeman, Ronald A.

; APPLICANT: Chisholm, George

; TITLE OF INVENTION: NOVEL METHODS FOR THE PRODUCTION OF GELATIN AND

; TITLE OF INVENTION: FULL-LENGTH TRIPLE HELICAL COLLAGEN IN RECOMBINANT

; TITLE OF INVENTION: CELLS

; FILE REFERENCE: 225002030400

; CURRENT APPLICATION NUMBER: US/09/585,887

; CURRENT FILING DATE: 2000-05-31

; PRIOR APPLICATION NUMBER: 09/289,578

; PRIOR FILING DATE: 1999-04-09

; PRIOR APPLICATION NUMBER: 60/084,828

; PRIOR FILING DATE: 1998-05-08

; NUMBER OF SEQ ID NOS: 11

; SOFTWARE: Patent in Ver. 2.0

; SEQ ID NO 9

; LENGTH: 1461

; TYPE: PRT

; ORGANISM: Homo sapiens

US-09-585-887-9

Query Match 96.6%; Score 2966; DB 4; Length 1461;

Best Local Similarity 95.6%; Pred. No. 8.9e-177;

Mismatches 520; Conservative 21; Mismatches 3; Indels 0; Gaps 0;

QY	1	GSEGGVGRGPPGPPAGAAGCPGADGCEPGAKGADGAPGIAGAPGCPGARGPSGPE	60
DB	359	GSEGGVGRGPPGPPAGAAGCPGADGCEPGAKGADGAPGIAGAPGCPGARGPSGPE	418
QY	61	GPGGPPGPKGDSGEPGAPGSKGDTGAKGEPGVGVGPPGPGAGBEKPKGARGEPGTGLP	120
DB	419	GPGGPPGPKGDSGEPGAPGSKGDTGAKGEPGVGVGPPGPGAGBEKPKGARGEPGTGLP	478
QY	121	GPGRGPGSGRGFGADGAVAGPKPAGRSFGPAGPKSGFGAGRGCEAGLPGAKGLT	180
DB	479	GPGRGPGSGRGFGADGAVAGPKPAGRSFGPAGPKSGFGAGRGCEAGLPGAKGLT	538
QY	181	GSFGSGPPDKTKGPPGAGEDGCPGPPGARGAGVWGFPGPKGAAGEPKGAGRGVVP	240
DB	539	GSFGSGPPDKTKGPPGAGEDGCPGPPGARGAGVWGFPGPKGAAGEPKGAGRGVVP	598
QY	241	GPFGVAGPAGKDGAGAEFGPPGAPGARGEGPAGSGFFGLPGPAPGPPGAEAKPGBE	300

Db 441 GPPGAVGPAGKDGEAGAQGP GPPAGPAGERGEQGPAGSPGFQGLPGPAGPPGEAGKPGEQ 500

QY 334 GPGADGAPGDDGAKGADAGAPGAGSEAGLPGKGDGADGAGPKGAD 393
 Db 571 GPRGANGAPGNDGAKGADAGAPGAGSQAAGLPGKGDGADGAGPKGAD 630
 QY 394 GSPGKDVRLGTGPIGPPGAPAGDKGSGSPGAPGTCARGAPDRGEPGPPGAPGA 453
 Db 631 GSPGKDVRLGTGPIGPPGAPAGDKGSGSPGAPGTCARGAPDRGEPGPPGAPGA 690
 QY 454 GPGADGEPGAKGEPGADGAKGADGAGPPGAPGAPGPIGDUVAPGAKGARGSAGPPGAT 513
 Db 691 GPGADGQPGAKGEPGADGAKGADGAGPPGAPGAPGPIGDUVAPGAKGARGSAGPPGAT 750
 QY 514 GFPGAAGRVGPPGSPGADGAGPPGPPGAPGKEG 544
 Db 751 GFPGAAGRVGPPGSPGADGAGPPGPPGAPGKEG 781

RESULT 9

US-09-548-608-18

; Sequence 18, Application US/09548608

; Patent No. 6355442

; GENERAL INFORMATION:

; APPLICANT: Bode, Martin

; APPLICANT: Oviatt, Per

; TITLE OF INVENTION: A Method for Assaying Collagen Fragments

; TITLE OF INVENTION: In Body Fluids, A Test Kit and Means for Carrying Out the

; TITLE OF INVENTION: Method and Use of the Method to Diagnose the Presence of

; TITLE OF INVENTION: Disorders Associated with the Metabolism of

; NUMBER OF SEQUENCES: 21

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Darby & Darby PC

; STREET: 805 Third Avenue

; CITY: New York

; STATE: New York

; COUNTRY: USA

; ZIP: 10022

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

; COMPUTER: IBM PC compatible

; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: Patentin Release #1.0, Version #1.25

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/09/548,608

; FILING DATE:

; CLASSIFICATION:

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: 08/187,319

; FILING DATE:

; ATTORNEY/AGENT INFORMATION:

; NAME: Gogoris, Adda C

; REGISTRATION NUMBER: 29,714

; REFERENCE/DOCKET NUMBER: 4305/08701

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: 212-527-7700

; TELEFAX: 212-753-6237

; TELEX: 236687

; INFORMATION FOR SEQ ID NO: 18:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 1341 amino acids

; TYPE: amino acid

; TOPOLOGY: linear

; MOLECULE TYPE: protein

; ORIGINAL SOURCE:

; ORGANISM: Homo sapiens

; IMMEDIATE SOURCE:

; CLONE: COLLAGEN ALPHA 1 (1)

US-09-548-608-18

Query Match

Best Local Similarity 90.9%; Score 2791.5; DB 4; Length 1341;

Matches 500; Conservative 21; Mismatches 23; Indels 27; Gaps 4;

QY 1 GSGGEGVRGEPGPPGAPGAGPAGDPGADG---FGAKGADGAPGAGPAGPPGAG--- 54

Db 211 GASGPMGPRGPPGPPGKGBBZAGKPGRCZRGPPFPZGARGLPGTAGLPGMKGHRGFS 270
 QY 55 -----GSPGPEG-----PG--GPPGPKGDSGEPGAPGSKGDTGAKGEPGV 93
 Db 271 GLBGAKBAGBAGPAGKZPGSPGZBGAFCZMGPPGPKGNSGEPGAPGSKGDTGAKGEPGV 330
 QY 94 GVEGPPGAPGEGKPGARGPPTGLPFPFCERGGPSRGGFGADGVAGKGPAGRGSP 153
 Db 331 GVQPPGAPGEGKRGARGPPTGLPFPFCERGGPSRGGFGADGVAGKGPAGRGSP 390
 QY 154 GPAGPKSPGAPGPGAGLPGAKLGTSGSPGPDCTGPPGPGAGEDGPPGPPGAR 213
 Db 391 GPAGPKSPGAPGPGAGLPGAKLGTSGSPGPDCTGPPGPGAGEDGPPGPPGAR 450
 QY 214 GEAGVMGPPGPKGAAGPFGKAGRGVFPFGCAVGPAGKDGGAEGPPGAPGAGEE 273
 Db 451 GQAGVMGPPGPKGAAGPFGKAGRGVFPFGCAVGPAGKDGGAEGPPGAPGAGEE 510
 QY 274 GPAGSPGEGLPAGPPGAPGKPGEBGVFDICAPGSPGARGPFGPPGREGVGP 333
 Db 511 GPAGSPGEGLPAGPPGAPGKPGEBGVFDICAPGSPGARGPFGPPGREGVGP 570
 QY 334 GPPGADGAPGDDGAKGADGAPGAPGSEGAPGLEGMPGCGAAGLPGPKGDRGADGPKGAD 393
 Db 571 GPRGANGAPGNDGAKGADGAPGAPGSCQAGLQGMFGERGAAGLPGKGDGADGPKGAD 630
 QY 394 GSPGKDVRLGTGPIGPPGAPGAPGDKGSGSPGAPGTCARGAPDRGEPGPPGAPGA 453
 Db 631 GSPGKDVRLGTGPIGPPGAPGAPGDKGSGSPGAPGTCARGAPDRGEPGPPGAPGA 690
 QY 454 GPGADGEPGAKGEPGADGAKGADGAGPPGAPGAPGPIGDUVAPGAKGARGSAGPPGAT 513
 Db 691 GPGADGQPGAKGEPGADGAKGADGAGPPGAPGAPGPIGDUVAPGAKGARGSAGPPGAT 750
 QY 514 GFPGAAGRVGPPGSPGADGAGPPGPPGAPGKEG 544
 Db 751 GFPGAAGRVGPPGSPGADGAGPPGPPGAPGKEG 781

RESULT 10

US-08-468-996-10

; Sequence 10, Application US/08468996

; Patent No. 6645504

; GENERAL INFORMATION:

; APPLICANT: Weiner, Howard

; APPLICANT: Miller, Ariel

; APPLICANT: Zheng, Zheng

; APPLICANT: Ahmad, Al-Sabbagh

; TITLE OF INVENTION: BYSTANDER

; TITLE OF INVENTION: GLUCAGON

; FILE REFERENCE: 1010/16959-US3

; CURRENT APPLICATION NUMBER: US/08/468,996

; CURRENT FILING DATE: 2003-02-07

; PRIOR APPLICATION NUMBER: US 07/843,752

; PRIOR FILING DATE: 1992-02-28

; PRIOR APPLICATION NUMBER: US 07/460,852

; PRIOR FILING DATE: 1990-02-21

; PRIOR APPLICATION NUMBER: US 07/596,936

; PRIOR FILING DATE: 1990-10-15

; PRIOR APPLICATION NUMBER: US 07/065,734

; PRIOR FILING DATE: 1987-06-24

; PRIOR APPLICATION NUMBER: US 07/454,486

; PRIOR FILING DATE: 1989-12-20

; PRIOR APPLICATION NUMBER: US 07/487,732

; PRIOR FILING DATE: 1990-03-02

; PRIOR APPLICATION NUMBER: US 07/551,632

; PRIOR FILING DATE: 1990-07-10

; PRIOR APPLICATION NUMBER: US 07/379,778

; PRIOR FILING DATE: 1989-07-14

; PRIOR APPLICATION NUMBER: US 07/607,826

; PRIOR FILING DATE: 1990-10-31

; PRIOR APPLICATION NUMBER: US 07/595,468

SUPPRESSION OF TYPE 1 DIABETES BY ORAL ADMINISTRATION OF:

; PRIOR FILING DATE: 1990-10-10
 ; NUMBER OF SEQ ID NOS: 13
 ; SOFTWARE: Patent in version 3.1
 ; SEQ ID NO 10
 ; LENGTH: 1017
 ; TYPE: PRT
 ; ORGANISM: Homo sapiens
 US-08-468-996-10

Query Match 72.5%; Score 2226; DB 4; Length 1017;
 Best Local Similarity 72.6%; Pred. No. 5.2e-131;
 Matches 395; Conservative 39; Mismatches 110; Indels 0; Gaps 0;

QY	1	GSEGGEGVGRGPPGPPGAGAGPAGDPCADGECRAGKAGDAGAGIACAGPFCGARGSPGE	60
DB	184	GREGAGQGRGPPCTFGSPCPAGASGNPTDGIKAKGSAGAFIAGAPFFGFRGPPDPQ	243
QY	61	GGGPPGPKGDSGEFAPSGKXGDTCAKGEPPVGVGPPGPPAGECKPGARGEPPGPTGLP	120
DB	244	GATGLPGKGTQKFGIAFGXGEQPKGEPGAPGQAGPAGCEKRGARGEPGGVPI	303
QY	121	GGPGERGGSRGPPGADGVAGPKPAGERGSCPPAGPKGSPCEAGRCPCEAGLPGAKGLT	180
DB	304	GPFGERGAPGNRFFGQDGLAGPKGAPCERGPSGLAGFKGANGDPPRPGPLPGARGLT	363
QY	181	GSFSGPPGDKTGPDPGAGEDCRPCPPGPPCARGECAGVMGFPFGKGAAGEPGKAGRGVP	240
DB	364	GRFGDAGQGVKVPSSGAFGEDRGPFGPPGQARGQPGVMGFPFGKANGCEPGKAGEKGLP	423
QY	241	GGPGAVGPAKDGAGAGRPDPGAPGAGERGEGPAGSPGFEGLPGAPGPPGEAKPGEE	300
DB	424	GAPGLRGLPGKDGETGAEGLPPGAPGAGERGEQAGPSPGFGQLPGPPPGGEAGKPGDQ	483
QY	301	GVPGDLGAPGPGSARGEGPFGPGRGVGPPGPPGAPGADCAPDDGAKGDAGAPGAPGE	360
DB	484	GVPGAGAPCLVPRGERGFPGERGSPGAQGLQGPRGLPCTPTDGPKASGAPFGPAQ	543
QY	361	GAPGLEMPGERGAAGLPKPKDRGDAGPKADGSPGKGVRLGTGPIGPPGAPAGDVK	420
DB	544	GPPLQGMGERGAAGIAGPKDGRDVGKGEKPEGAPGKDGARGLTGPIGPPGAPAGGK	603
QY	421	GESGSGPAGPTGARGAPGDRGERGPPGPPAGFAGPPGADGCEPKAGEPGDAGAKGDGPP	480
DB	604	GEVGGPPGASGAGAPGERGTGPPGPPAGFAGPPGADGQPKAGEQGEACQKGDAGAP	663
QY	481	GRAGPAGPPGPIGDVAFGAPKAGSAGSPGPGATGPPGAAGRVGPPGSGDAGPPGPPA	540
DB	664	GPQPGSAGPPGPGTGVTPGPKARGAQQPPGATGPPGAAGRVGPPGSGNGNPPGPPGPP	723
QY	541	GKEG 544	
DB	724	GKDG 727	

RESULT 11
 US-08-931-820-3
 ; Sequence 3, Application US/08931820
 ; Patent No. 6010863
 ; GENERAL INFORMATION:
 ; APPLICANT:
 ; TITLE OF INVENTION: Assay for collagen degradation
 ; NUMBER OF SEQUENCES: 4
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: Floppy disk
 ; COMPUTER: IBM PC compatible
 ; OPERATING SYSTEM: PC-DOS/MS-DOS
 ; SOFTWARE: Patent in Release #1.0, Version #1.25 (EPO)
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: US/08/931,820
 ; FILING DATE:
 ; CLASSIFICATION: 435
 ; PRIOR APPLICATION DATA:
 ; APPLICATION NUMBER: EP 96202596.1

